

# **Monopoly Bidding Strategy Under the Ontario MPMA and System Reliability**

by

**Aiming He**

**B.S. Xiangtan Mining Institute**

**M.A. Xiamen University, M.A. Lakehead University**

A thesis submitted to the Faculty of Graduate Studies  
and Research in partial fulfilment of the requirement for the degree of  
Doctor of Philosophy  
in the  
Department of Economics

CARLETON UNIVERSITY

Ottawa, Ontario

March 2007

Copyright © 2007, Aiming He

All rights reserved. This work may not be reproduced  
in whole or in part, by photocopying or other means without  
permission of the author



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
*ISBN: 978-0-494-27100-1*  
*Our file* *Notre référence*  
*ISBN: 978-0-494-27100-1*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

## Abstract

This thesis explores two aspects of the economics of electric power systems. The first is the impact of the Market Power Mitigation Agreement (MPMA) which was in place in Ontario over the period 2002-2005 on the offer strategy of a hypothetical monopoly generator. The second is the relationship between engineering and economic models of power system reliability.

Chapter 1 briefly reviews the history and potential problems of electricity deregulation in Ontario and puts forth the purposes of this thesis.

Chapter 2 analyzes the welfare implications of the MPMA in the context of a single period monopoly model. The main findings in this chapter are: (1) that the MPMA is welfare improving; (2) that there are multiple CRQ's that can lead to the social optimum; and (3) that the second best MPMA requires a higher CRQ and thus a lower price cap under demand uncertainty than the expected break-even levels.

Chapter 3 constructs a two period model that assumes two types of generation, thermal and a limited hydro generation capability that can be shifted frictionlessly between periods. In this chapter, the main conclusions are: (1) that peak consumers benefit more from the MPMA than off-peak consumers; (2) that the hydro resource is more efficiently used under the MPMA than under an unregulated monopoly; and (3) that marginal cost pricing is not socially optimal.

Chapter 4 constructs a linkage between the demand for electricity and system reliability by deriving an endogenous demand function for operating reserve based on a statistical model. In this chapter, the main findings are: (1) that the traditional OR requirement may be greater or smaller than the optimal reliability level; (2) that a monopoly generator will produce more electricity if the demand for OR is statistically determined; (3) that if the socially optimal MPMA is applied to the energy market only, the energy market is socially optimal but the OR supply is too low; and (4) that capacity subsidies or capacity market can be welfare improving.

Chapter 5 summarizes the main conclusions reached in Chapters 2 to 4.

# Acknowledgements

I would like to express my deep appreciation to my supervisor Professor Don McFetridge. He has not only provided me guidance on and valuable inputs into this thesis, but also led me into the electricity industry to which I have contributed several years and likely many years ahead. The experience of learning from and working with him will benefit me for a life time.

I also would like to thank Professor Zhiqi Chen, Professor Gamal Atallah, and Professor Christopher Worswick. During the past several years, they have provided me useful comments on my thesis. Without their support, I wouldn't have had the opportunity to fulfil my studies.

I am also grateful for my co-workers Brian Rivard, Jack Lubek, and manager Richard Penn at the Independent Electricity Operator. They have spent plenty of time and effort in past years to help me understand the function of the electricity market in Ontario as well as in other countries, and provided many invaluable suggestions for this thesis.

I would extend my thanks to other economics professors at Carleton University and the University of Ottawa for their teaching me so many valuable courses, and Ms. Marge Brooks for her timely communications when I was working off campus.

Last but most important, I am grateful to my dear wife Lanfang for her unconditional love, full support and understanding. She has taken most if not full responsibilities in past years to raise our two lovely sons, Richard and Mark.

All errors remain mine.

# Table of Contents

Abstract .....	ii
Acknowledgements .....	iii
Table of contents .....	iv
List of Tables .....	vii
List of figures .....	viii
Chapter 1 Introduction .....	1
1.1 Market Evolution .....	2
1.2 Market Structure .....	6
1.3 Potential Problems with the Ontario Market .....	9
1.4 The Future .....	13
References .....	14
Chapter 2 A One-Period Model .....	15
2.1 Introduction .....	15
2.2 Review of Regulation Theories and Practices .....	17
2.3 Market Power Mitigation Agreement .....	20
2.4 Model .....	27
2.5 Welfare Impact of the MPMA .....	30
2.6 Impact of the Price Spike Adjustment .....	34

2.7 Multiplicity of Optimal Policies .....	36
2.8 Demand Uncertainty .....	39
2.9 Second Best Rebate Mechanism Under Demand Uncertainty .....	45
2.10 Supply Uncertainty 1: Fuel Cost .....	48
2.11 Supply Uncertainty 2: Forced Outage .....	50
2.12 Conclusions .....	54
References	56
Chapter 3 A Two-Period Model .....	58
3.1 Model .....	60
3.2 Non-uniqueness of Policies to Inducing a Monopolist to Bid at a Price Cap .....	74
3.3 Welfare Implication of the MPMA.....	83
3.4 Water Allocation .....	84
3.5 Social Optimum.. ..	85
3.6 Impact of the Water Stock .....	89
3.7 The Impact of Demand Shocks .....	91
3.8 Imports .....	105
3.9 Marginal Cost Pricing .....	113
3.10 Conclusions .....	117
References .....	119
Chapter 4 Reliability and the MPMA .....	121
4.1 Introduction .....	121

4.2 Approaches to Calculating Operating Reserve Requirements .....	125
4.3 A Brief Review of Literature on Pricing Operating Reserve .....	129
4.4 A Demand Function for Operating Reserve .....	133
4.5 The Model .....	139
4.6 A Capacity Constrained Monopolist .....	148
4.7 First Best and Second Best .....	151
4.8 The Optimal MPMA .....	155
4.9 A Decentralized Market .....	160
4.10 Capacity Allocation in Shortage: Monopoly vs. Competition .....	171
4.11 Policies Leading to the Social Optimum.....	175
4.12 Conclusions .....	177
References .....	180
Chapter 5 Conclusions .....	183
<b>Glossary</b> .....	191

## List of Tables

Table 3.1: Comparison of an Unconstrained Monopolist and A Monopolist Who Bids above the Price Cap.....	68
Table 3.2: Comparison of a Monopolist Who Bids above the Price Cap and a Monopolist Who Bids below the Price Cap .....	71
Table 4.1: Operating Reserve and Reliability Levels .....	129
Table 4.2: Demand for Operating Reserves .....	136

## List of Figures

Figure 2.1: Equilibrium under the MPMA .....	29
Figure 2.2: High or Low Cost Equilibria under the MPMA .....	30
Figure 2.3: Welfare Improvement under the MPMA .....	31
Figure 2.4: Failure of the MPMA .....	33
Figure 2.5: Effect of the Price Spike Adjustment .....	35
Figure 2.6: Profit as a Function of the Price Cap .....	39
Figure 2.7: Demand Shocks and the Optimum .....	40
Figure 2.8: Equilibrium with a Supply Shock .....	50
Figure 2.9: Imports with a Forced Outage .....	54
Example Figure 3.1: Conditions of CRQs Given a Price Cap .....	72
Figure 3.1: Thermal Production in Equilibrium under the MPMA .....	73
Figure 3.2: Illustration of Monopoly Equilibrium (Bidding below the Price Cap) ...	76
Figure 3.3: Illustration of Monopoly Equilibrium (Bidding above the Price Cap) ...	78
Figure 3.4: Illustration of Monopoly Equilibrium (Bidding at the Price Cap) .....	79
Figure 3.5: Optimal Bidding Strategy .....	83
Figure 3.6: Welfare Maximization Equilibrium .....	89
Figure 3.7: Welfare Maximization when Water is Insufficient for Both Periods .....	91
Figure 3.8: Thermal Production with a demand Shock (Scenario 1) .....	95
Figure 3.9: Thermal Production with a demand Shock (Scenario 2) .....	101
Figure 3.10: Profitability of Spilling Water .....	115
Figure 4.1: Operating Reserve demand Curves .....	136

Figure 4.2: Demand for OR is Proportional to the Demand for Energy .....	139
Figure 4.3: The Difference between the Social Optimum and Competitive Equilibrium .....	165
Figure 4.4: Equilibrium Energy and OR prices .....	173

## Chapter 1 Introduction

The Ontario wholesale electricity market was launched on May 1, 2002, after 6 years' consultation, designing, legislation, and testing of software. The purpose of restructuring this traditional monopoly industry into a competitive market was to induce more efficient production and investment decisions based on market signals. The market has changed dramatically since its inception, especially –since the passage of the *Electricity Restructuring Act 2004* in December, 2004. Long-term contracts procured by the government through its agency -- Ontario Power Authority -- have largely replaced the role of the spot market in guiding investment decisions. In the coming ten to twenty years when these contracts are still in effect, the role of the spot market will be limited to ensuring that dispatch is efficient and competitively priced.

When I started my thesis, the Ontario wholesale electricity market was just launched and relatively little had been written on it. The intent of this thesis was to conduct a comprehensive analysis on the impact of the Market Power Mitigation Agreement (MPMA) signed between the Ontario government and Ontario Power Generation Inc. (OPG) on the offer strategy of OPG and to compare the MPMA with an optimal regulatory scheme. However, the market evolved so fast that the MPMA was supplanted by new regulations, especially Bill 100, and the core mechanism in the MPMA now applies to a relatively small portion of Ontario generating capacity. As a consequence, the analysis of the MPMA, while retaining its theoretical implications, offers somewhat less in the way of immediate policy implications.

This thesis is constructed as follows. This introduction will briefly discuss the history of the Ontario wholesale electricity market, as well as its structure and problems. Chapter 2 introduces a simple one period model under a monopoly setting, assesses the effect of the MPMA on a monopoly generator's bidding strategy, and derives the optimal policy. Chapter 3 models a monopoly generator in a two period setting, assesses the impact of the MPMA on hydroelectric resources and also derives the optimal policy. Chapter 4 extends the energy-only market to include an operating reserve (OR) market, shows the possibility of constructing an OR demand curve based on probability of machine failure, derives the optimal policy for a monopolist and shows policy implications in a competitive environment. The final chapter will summarize my findings in this thesis.

### *1.1 Market Evolution<sup>1</sup>*

Traditionally, Ontario's electricity industry was dominated by Ontario Hydro, a vertically integrated monopoly, which was solely owned by the Ontario government and regulated by the Ontario Energy Board. The majority of consumers purchased electricity from Ontario Hydro through their local municipal utility company at a fixed price that bundled together generation, transmission, distribution and retail service.

In 1995, in response to the mounting debt of Ontario Hydro<sup>2</sup> and inspired by both the successful electricity market in the United Kingdom and serious discussions in favour of

---

<sup>1</sup> For more details on the restructuring of Ontario electricity industry between 2002 and 2003, interested readers may refer to Trebilcock and Hrab (2005).

<sup>2</sup> At the time when the restructuring started in 1998, Ontario Hydro had a debt of \$38 billion in comparison to an asset value of \$19 billion, most of which resulted from the expansion of generation capacity and cost overruns in nuclear station construction (especially the Darlington station).

market solutions in the United States, the Ontario government appointed a committee led by Donald Macdonald to explore the possibility of a competitive electricity market. The committee filed a report to the government in 1996, making several recommendations to restructure the province's electricity industry, including breaking up Ontario Hydro, privatizing some generation and transmission assets and forming a competitive electricity market. Following the report, the provincial government released a White Paper in 1997, outlining the direction of industry restructuring, and created the Market Design Committee in January 1998.

Following the Market Design Committee's recommendations, the provincial government passed its *Energy Competition Act* (Bill 35) in October 1998, setting out the framework for a competitive electricity market. Ontario Hydro was then split into five companies: Ontario Power Generation Inc. (OPG), Hydro One, the Independent Market Operator (IMO which was later renamed as the Independent Electricity System Operator—IESO), the Ontario Electricity Financial Corporation (OEFC), and the Electricity Safety Authority (ESA). OPG acquired all of Ontario Hydro's generation capacity but was subject to the restrictions of the MPMA which was intended to constrain its market power. Hydro One took on all transmission assets and distribution responsibilities from Ontario Hydro and is subject to full regulation. The IESO is responsible for running the wholesale electricity market and ensuring the reliability of the electricity system. The OEFC services and pays down the stranded debt of the former Ontario Hydro. The ESA sets safety standard for wiring installations and is responsible for equipment and appliance certification.

To prevent OPG from abusing its dominant market position, the OEB licence granted to OPG included the MPMA, which mandated that OPG be subject to a price cap (\$38/MWh) above which OPG had to pay a rebate to consumers on 90% of its forecast domestic sales. The MPMA also required OPG to divest its generation capacity over a prescribed timeline<sup>3</sup>.

In the early stages of market restructuring, it seemed that the Ontario government intended to privatize the transmission company Hydro One, but was unsuccessful after a court ruling<sup>4</sup>. Hydro One is still 100% under government control.

In light of the requirements and incentives of the MPMA, OPG showed strong interest in selling or transferring the control of some generation capacity both before and after the wholesale market was launched. Indeed, it leased out the Bruce nuclear stations to British Energy in May 2001, sold the Mississauga River hydro plants to Brascan in March 2002, and put up two fossil-fuelled generation stations for sale. However, because the offered sale price was too low and potential buyers refused to convert them into gas fired generation, the deal was pulled out off the table by the government (Power Week Canada, November 12, 2002).

As the wholesale electricity price started to skyrocket in the first summer after market opening (the average price jumped from \$29/MWh in May to \$75/MWh in September, 2002), the electricity market became a headline in the media and a political topic in the

---

<sup>3</sup> See Chapter 2 for details.

<sup>4</sup> On April 19, 2002, Mr. Justice Arthur Gans ruled that the Electricity Act was “not intended to deal with privatization” after two unions challenged the government’s decision to sell Hydro One.

provincial parliament. In response to increasing criticism of the high price, the Progressive Conservative government passed a new electricity law – *Electricity Pricing, Conservation and Supply Act 2002* on December 9, 2002, which froze the retail price to small consumers until at least 2006.

After the Liberal Party won the provincial election in 2003, *the Electricity Restructuring Act 2004* (Bill 100) was passed in December 2004, which introduced more regulations on OPG and created a new agency, – the Ontario Power Authority, to sign and oversee the long term contracts with private suppliers. Under the new law and its related by-laws, OPG assets were separated into two categories: prescribed and non-prescribed assets. The prescribed assets include all nuclear stations and base-load hydroelectric stations; each type is subject to a different contract price. The output by prescribed assets accounted for 45 percent of total Ontario production and 70 percent of OPG’s total output in 2004. The non-prescribed assets consist of all fossil (coal or gas-fired) generation capacity and all peaking hydroelectric stations, accounting for 25 percent of Ontario energy and 30 percent of OPG’s total output. All non-prescribed assets, except the Lennox station, are subject to a price cap and rebate mechanism similar to that in the MPMA but with a higher price cap \$47 in 2005 and 85% of total output on non-prescribed assets. As a result, the output that is subject to price cap regulation is much smaller than before.

The new legislation had a significant impact on the wholesale electricity market. First, the volume that is exposed to the real time price significantly dropped, given that most of OPG’s production is under contract. According the MSP, only about 20% of delivered

power is now settled in the spot market (MSP, December 2005). Second, the spot price is no longer the only signal for potential investments. In fact, a potential entrant probably wouldn't come to the market unless he/she has a deal with OPA, given that OPA has signed many bilateral contracts with new suppliers through the Request for Proposals (RFP) process and plans for more. Third, OPG has less incentive to exercise its market power given that most of its output is now sold under fixed price contracts rather than at the real time (spot market) price.

### *1.2 Market Structure*

The Ontario electricity market consists of different types of participants: generators, transmitters, loads (large consumers who are directly connected to the IESO-administrated grid), importers/exporters, wholesalers, distributors, retailers, etc. Small consumers such as households do not directly participate in the market but purchase electricity from their local distribution company or retailers. Currently, there are 34 generation companies with a total of 262 generation units. The maximum generation capacity in Ontario is (40,000) MW (of which a large part is hydroelectric capacity which cannot produce at full capacity in most times due to energy limits). Transmitters don't trade in the market but are responsible for delivering power from generators or importers to consumers or exporters. Directly connected wholesale loads include more than 90 large industrial and commercial consumers, who account for about 15 percent of total Ontario demand. There are about 60 registered intertie traders who arbitrage the price differences between Ontario and neighbouring markets. Distributors include the

municipal utilities and Hydro One's distribution arm. They draw power from the IESO-administrated grid and deliver it to local residential, commercial and industrial consumers and retailers.

On the demand side, residential consumers account for about 33.3% of total consumption, Commercial consumers 37.1%, and industrial consumers 29.3%.

Residential consumers and most commercial and industrial consumers are not responsive to the spot market price because they face a fixed regulated price and lack real time meters to record their real time consumption. With the provincial government mandating the installation of "smart meters" which can record real time consumption and the implementation of a real time pricing or time of use pricing regime, it is expected that the peak consumption by these "non-responsive" consumers will drop significantly. At present, only a few large industrial consumers totalling 2500 MW are actively participating in the market and following the IESO's dispatch instructions that are issued every five minutes.

On the generation side, there are two types of generators in the wholesale market: dispatchable and non-dispatchable. Dispatchable generators must be able to adjust output in response to dispatch instructions of the IESO. This group accounts for more than 90% of total output. Non-dispatchable generators (so called Non-utility Generators, or NUG's) are self-scheduling or intermittent generators<sup>5</sup> and do not submit offers to the market. Instead, they submit estimates or forecasts of energy production independent of

---

<sup>5</sup> An intermittent source such as a windmill generator has difficulties submitting an offer or following dispatch since its output depends almost solely on the wind's strength.

the market price. Most of these generators have a heritage contract<sup>6</sup> with the OEFC (succeeded from Ontario Hydro) at a fixed price, and are thus insulated from real time price impacts. There is another group of generators called embedded or distributed generators, who are not directly linked to the IESO-administrated grid but generate power for the local distribution grid or for the owner's self-consumption.

Generators can be split into four categories, according to generation technology or fuel type: nuclear, fossil (including coal and gas/oil), water, and waste. Nuclear generation capacity uses uranium as its fuel and accounts for about 50% of total energy production in Ontario. Fossil-fuel generators consume either coal, or natural gas, or oil, accounting for about 25% of market share. Hydroelectric generators use water flow to generate power, accounting for 25% of market share, and some small generators can generate electricity by combusting wastes, such as wood, garbage, or using residual heat, etc., for the remaining market share.

The Ontario wholesale electricity market is structured with three product / sub-markets: the real time markets (i.e. spot markets), financial markets, and procurement markets. The real time markets include the energy market and the operating reserve market. The energy market trades electrical energy. The operating reserve market trades stand-by capacity that is ready to respond to any significant generating machine failure or transmission line failure. Financial markets deal with transmission rights, allowing the intertie traders to hedge the risk of transmission congestion on the interties. Procurement markets enable the IESO to ensure the reliability of Ontario's power system by acquiring

---

<sup>6</sup> The contracts were signed between the NUG's and Ontario Hydro long before the market opening.

products such as black-start capability<sup>7</sup>, Reliability Must Run capability<sup>8</sup>, and voltage support or frequency maintenance service.

### *1.3 Potential Problems with the Ontario Market*

The first problem with the Ontario market is the existence of market power which may increase financial risk for a new entrant and result in supra-competitive pricing. OPG which took over Ontario Hydro's generation assets and has divested some of them during the course of market restructuring, accounts for 70% of the market. The second largest producer is Bruce Power, who operates Bruce nuclear stations and accounts for 21% of market share. Because production of nuclear power is not flexible<sup>9</sup>, Bruce Power can be regarded as a price taker in the market. Brascan is the third largest generator, who owns many small hydro stations and has a market share of 1 percent. TransAlta Energy Corporation is the fourth largest, accounting for another 1 per cent of market share. Although OPG was subject to the MPMA regulation and is now under contract with respect to most of its production, it is owned solely by the Ontario government and thus may serve other objectives on behalf of the government. This may create market risk for other producers and new entrants.

---

<sup>7</sup> Black-start capability is a generator's ability to help restore the province's power system after a blackout without relying on an external supply of electricity.

<sup>8</sup> A peaking generator may be unable to survive under prevailing market prices, and thus require a Reliability Must Run contract with the system operator for compensation of staying in service.

<sup>9</sup> Nuclear generation has limited ability to respond to the system operator's instruction to quickly ramp up or down due to technological difficulty or safety concerns. Furthermore a nuclear station has a much longer shut down time: it cannot be started within two days after it is shutdown.

OPG's divestiture process was terminated under the 2004 legislation. Although most OPG capacity is subject to either a contract price or a price threshold, many potential entrants are still concerned about the market power of OPG, as well as other political uncertainty which may affect market prices. As the Toronto based Energy Probe and CIBC World Markets observed, the low price in summer 2003 might have been a result of a change in OPG's bidding strategy which served political objectives. As a result, there is a reluctance to invest in new generation. The consequence is that under the current market situation, the market price is a highly uncertain signal of investment opportunities. The long term contracts signed with the government have worked as a risk free deal, in which the profit of a new entrant is not affected by market outcomes. It is expected that a new entrant will not enter the market unless it has such a contract with the government.

The second problem is the uniform price system. Unlike the electricity markets in the United States that typically take what is known as the Locational Marginal Pricing approach<sup>10</sup>, the Ontario market takes a uniform price approach, which requires all consumers to pay the same price regardless of transmission constraints. Thus the IESO, the system and market operator, has to run two separate sequences: one to determine the market clearing price, and the other to determine the electricity production and flows based on transmission limits, generator characteristics, and other constraints. This approach causes two serious problems. First, the cost of transmission congestion is masked because the uniform price is a fiction in many ways and rarely reflects the

---

<sup>10</sup> Based on transmission line congestion, a market may be divided into many zones with each having a locational price. When the transmission line is congested between two adjacent zones, the zone with import congestion will have a higher price. Otherwise the two zones may have an identical price.

tightness of supply and demand conditions. For example, consumers in a load pocket (i.e. an area with import lines frequently congested) do not have to pay high prices although high cost generation in the area is paid to generate power for them<sup>11</sup>. This approach leads to subsidies from consumers in low cost areas to those in high cost areas. Thus the uniform price results in inefficient consumption. Second, because the price doesn't reflect the true supply/demand conditions, a new entrant either has no information on where the most needed place lies or has no incentive to build a plant in the most needed place. Similarly a generator has no incentives to avoid an area which already has a surplus of generation. As a result, the uniform price system may lead to inefficient investment. Other problems include inefficient export and bidding strategy distortion due to the inverse incentive of the Congestion Management Settlement Credit (MSP, December 2005, June 2006). The Market Surveillance Panel is in favour of a locational pricing model, or at least kind of zonal pricing (MSP, June 2006).

The third problem is the limitation of Ontario's market design to a spot market only. Unlike many markets around the world, Ontario wholesale electricity market has neither a Day Ahead Market (DAM) nor a capacity market. While it is true that production efficiency occurs in the real time market, a DAM has many benefits. Firstly, the existence of a DAM discourages exercise of market power (Wolak, 1999). The intuition is that a generator who sells most of its energy in a DAM has less incentive to withhold its capacity in real time because the energy sold in a DAM at the DAM price cannot

---

<sup>11</sup> To compensate those generators who are constrained on or off (a generator is constrained on if its offer is higher than the market price but it is needed for reliability, and a generator is constrained off if its offer is lower than the market price but is not required to produce power), there is another side payment mechanism called Congestion Management Settlement Credit which pays those constrained on generators up to their offers and those constrained off generator the difference between the market price and their offers.

benefit from a higher price in real time. Secondly, the DAM price is a forecast of the real time price. With a longer time horizon, consumers, especially large consumers, can take steps to better respond to price signals. Thirdly, a load or a generator that has locked in the DAM price is immune to large swings in the real time price given that real time price is usually much more volatile than the DAM price. Through the DAM, a participant can thus hedge against price risk. Fourthly, a DAM results in a binding commitment for generators and intertie traders, and thus gives the system operator more assurance of performance, which helps to maintain the reliability of the power system. A capacity market itself has no intrinsic value (Oren, 2005), but can be considered as a means to remedy the ‘missing money’ problem associated with an energy only market (Joskow, 2006).

The fourth problem is the political uncertainty associated with the electricity market. At the beginning, the government promised a free market and was reluctant to intervene even when the price became higher and higher. As the criticism increased and the election date approached, the government froze the retail price for the majority of consumers. This action was a big blow to the market and pushed away potential investors since more regulations on the wholesale market could be foreseen, once the retail price was frozen. After the Liberal Party won the election in 2003, the course of deregulation was significantly changed. With Bill 100 passed in December 2004, regulation of OPG was fundamentally altered and new entry occurred only when the entrant had a long term contract with OPA. As coal-fired generation stations are phased

out and more contracted capacity comes into the market in coming years<sup>12</sup>, the spot market will have a smaller and smaller impact on generation payments and a smaller role in signalling efficient investment.

#### *1.4 The Future*

After more than four years of operation, Ontario's wholesale electricity market is still in its infant stage. However, the Ontario market is gradually taking root and there is no way going back to the old regulated world, despite the bumps in the road getting where it is today. As more private contracted capacity comes on-line in coming years, the electricity price will be much more competitive and stable than what we have seen in past three years, and the pressures on government will be lower. As a result, the role of the wholesale market is likely to become more and more important.

---

<sup>12</sup> The provincial government had initially vowed to shut down all coal units by 2007, but agreed OPG should keep some units as reserve until enough new capacity comes on-line.

## References

1. Joskow, P.L. (2006) “Competitive Electricity Markets and Investment in New Generating Capacity”, AEI-Brookings Joint Center for Regulatory Studies, Working Paper 06-14.
2. Market Surveillance Panel, Market Monitoring Report December 2005, June 2006, [www.oeb.gov.on.ca](http://www.oeb.gov.on.ca).
3. Oren, S. (2005) “Generation Adequacy Via Call Options: Safe Passage to the Promised Land”, University of California Energy Institute Working Paper WP-016.
4. Power Week Canada, November 12, 2002
5. Trebilcock, M. and R. Hrab (2005) “Electricity Restructuring in Ontario”, *The Energy Journal* 26(1): 123-146
6. Wolak, F. (1999) “An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market”, mimeo, Stanford University, Available at [www.stanford.edu/~wolak](http://www.stanford.edu/~wolak).

# Chapter 2 A One Period Model

## 2.1 Introduction

There are many approaches to mitigating market power or to discouraging exercise of market power in electricity markets. In most electricity markets in the United States, market power mitigation is automatic. The Automatic Mitigation Procedure (AMP) used by the system operators in New York, New England, California and Pennsylvania, New Jersey and Maryland (PJM) reduces a generator's bids if its bids fail three sequential tests<sup>13</sup>. Other mitigation approaches include compulsory contracts (in New York), licensing restrictions (in Britain), divesting large producers (in California), and a maximum bid cap (in all electricity markets, for example US\$1,000 in all U.S. markets, CDN\$2,000 in Ontario).

The approach taken in the Ontario wholesale electricity market is very different from those in the United States or other countries; it initially took the form of an agreement (Market Power Mitigation Agreement, or MPMA in short) between Ontario Power Generation (OPG), and the Ontario government. A key element in the MPMA was the price cap and rebate mechanism, which defines a threshold (\$38 per MWh), above which the OPG has to rebate to consumers most of the revenue that is earned above the cap. This was intended to reduce OPG's incentives to use its dominant position in the market.

---

<sup>13</sup> The three sequential tests are the Conduct Test, the Price Impact Test, and the Profitability Test. The Conduct Test tells whether bids are above reference bids which are based on a 90 day bid history. The Price Impact Test tells whether a bid price which fails the Conduct Test has a material price impact. The Profitability Test tests whether the increase in bid prices is profitable for the firm that exercises market power. If all the three tests are failed, the generator's bids are reduced to its reference bid level.

The MPMA rebate mechanism was significantly changed in March 2005 after the new provincial government passed the *Electricity Restructuring Act 2004* and its supplementary protocols. This thesis deals with the situation as it was between May 2002, when the market opened, and March 2005.

The MPMA has been criticized for its inadequacy in mitigating OPG's market power. Many market participants (including large consumer groups and smaller power producers) advocate that OPG should divest itself of most of its generating capacity thus obviating the need for the MPMA. Market opponents (including groups of small consumers, and one provincial political party) argued that the MPMA left too much profit to OPG and that more stringent regulations should be implemented. It seemed that both sides agreed that the MPMA left the OPG too much market power and too much incentive to increase the wholesale electricity price.

The main objective of this chapter is to analyze the effectiveness of the MPMA in constraining market power. The chapter will be structured as follows. First, we will briefly introduce regulation theory/practices and show how the price cap in the MPMA is different from other forms of price regulation. Then, we will explain the MPMA and its key components. In the remaining sections, we show how the MPMA would work in the context of a simple, single period monopoly model.

## 2.2 Review of Regulation Theories and Practices

In a modern regulatory system, there are generally two extreme types of regulatory schemes: profit-based regulation and performance-based (incentive) regulation. Under profit-based regulation, the firm is regulated to ensure a "fair" profit. This system is commonly known as the "rate of return" (RORR) or "cost plus" regulation. Under this scheme, regulators establish a permissible rate-of-return, which is sufficient to cover the company's costs, plus a "reasonable" profit that allows the industry to attract necessary capital investment. However this system has some major drawbacks as has been well-documented: (1) it doesn't induce cost minimizing behaviour (Averch and Johnson, 1962); (2) excessive risk is imposed on consumers (U.S. Department of Commerce, 1991, Joskow, 1974); (3) cost may be shifted from unregulated markets to regulated markets when the regulated firm is operating in both markets; and (4) the choice of operating technology may be inefficient (Braeutigam and Panzar, 1989). In fact, because economic profit is ensured by the "fair" rate-of-return on the capital investment, a firm subject to rate of return regulation can earn economic profits by over-investing in the capital stock relative to the profit maximizing level (Averch and Johnson, 1962) or more generally encouraging X-inefficiency (Joskow and Schmalensee, 1986). Averch and Johnson's critique of traditional regulation led to an examination of the incentive effects of regulation.

As an alternative to profit-based regulation, performance-based regulation has gained increased popularity in many regulated industries around the world. In the United States, performance-based regulation has been employed extensively in the telecommunications

industry for many years, and more than 40 states now employ this kind of regulation to monitor the intrastate operations of local exchange carriers.

Modern performance-based regulation differentiates between Bayesian and non-Bayesian mechanisms (Vogelsang, 2002). Bayesian mechanisms assume the regulator's lack of information takes the form of a subjective probability that the regulator holds about the parameters of the regulatory problem. In the asymmetric information regime, the effects of adverse selection (Baron and Myerson, 1982) and moral hazard (Laffont and Tirole, 1986, 1994) are added. A major result from the Bayesian literature is that regulated firms will make economic profits under the optimal mechanism. Non-Bayesian mechanisms attempt to use only observable and verifiable data and to be independent of a particular regulator. Because those data cannot be foreseen, these mechanisms are not optimal. The most popular performance-based approach, price cap regulation, is a blend between Bayesian and non-Bayesian mechanisms. The main Bayesian parameter is the *X*-factor, which has to be adjusted every few years and requires a judgement about the potential of the regulated firm to reduce cost.

As Sappington (2001) shows, performance-based regulation can be generally categorized into eight groups:

- (1) banded rate of return regulation --- the regulated firm is allowed to keep all earnings as long as the earnings fall within a pre-specified target range;
- (2) earnings sharing regulation (sliding scale regulation) --- the regulated firm and its consumers share all realized earnings;

- (3) revenue sharing regulation --- the firm and its consumers share revenues (rather than earnings) that exceeds certain level;
- (4) rate case moratoria --- the regulated firm agrees with the regulator to return projected earnings to target levels to avoid investigation;
- (5) price cap regulation --- a price cap is imposed on a certain product or service or a group of products or services;
- (6) partial deregulation --- the firm's products or service are classified into different categories (e.g. competitive, discretionary, and basic), and those considered to be competitive are removed from regulation;
- (7) yardstick regulation --- the allowed return to investment is based on its performance relative to the performance of comparable firms; and
- (8) options --- the regulated firm is allowed to choose one of proposed regulation plans.

Incentive regulation is not without drawbacks. For example, service quality may be reduced due to the strong incentives to minimize cost. Second, because risk is now significantly shifted to investors, the cost of capital is increased. Third, a rate shock (Isaac, 1991) when the price cap is reviewed for the following period may arise when the realized cost significantly increases in the previous period. And finally, the regulated firm may collect significant profit when prices significantly deviate from realized cost, which might be socially undesirable and politically unacceptable.

There exists one variant of price cap regulation, which is widely used in electricity markets ---soft price cap regulation. This is essentially not a price cap because it is “soft”, i.e. adjustable as long as the regulated firm or firms request a change and the regulator deems the change is “reasonable”. An example is the bid reference under the Automatic Mitigation Procedure (AMP) used in many U.S. electricity markets, which is based on 90 day historical bids, and can be violated if the offer submitter can provide “reasonable” explanations, for example a large increase in fuel cost. Similarly, the price cap (or more precisely price threshold) under the MPMA doesn’t limit OPG’s bid directly because OPG can bid at any level it thinks profitable.

### **2.3 The Market Power Mitigation Agreement**

Recognizing the dominant position of OPG, the Market Design Committee (MDC) established an agreement with OPG so as to constrain its market power, which later became the basis of the Order in Council issued by the Province of Ontario on March 24, 1999. This agreement was, in turn, based on the Market Power Mitigation Proposal that was unanimously endorsed by the MDC (reluctantly because actual break-up/divestiture was preferred, but giving up ownership completely was not an option allowed by the government).

The MPMA used the wholesale price of electricity as the major instrument for rewarding or discouraging the actions of OPG in moving towards its target market share as stipulated in the Agreement. Before the spot market opened, it seemed that OPG was

ready to divest capacity ahead of the schedule. It leased out the Bruce nuclear station to British Energy before market opening, and put up some fossil plants for sale, e.g. Lakeview, Lennox, Thunder Bay, Atikokan, and Mississagi hydroelectric units (Mississagi units were sold to Brascan Canada on May 17, 2002). After the market's opening in May 2002: "... little (divestiture) is occurring. Two coal-fired plants that had been floated for sale have been pulled off the table because of low offers and a potential buyer's unwillingness to convert the plants to natural gas." (Power Week Canada, November 12, 2002). After the 2003 provincial election, the divestiture agenda was totally withdrawn from the table since the provincial government introduced a new hybrid market mechanism<sup>14</sup> and determined to phase out all coal fired generation by 2007 (the date was further extended to 2014).

There were four key objectives in the Agreement:

- a. to discourage OPG from exercising market power from its dominant position in the period between market opening and the achievement of an effectively competitive market.
- b. to encourage actions by OPG to decontrol its generation capacity in specific segments of the supply schedule.
- c. to protect consumers from substantial increases in the wholesale price.
- d. to encourage new investment in the Ontario electricity market, or increase responsiveness of imports.

---

<sup>14</sup>To induce more competition, the McGuinty government issued two Requests for Proposals which invite new clean generation and renewable generation, besides the two-tier asset system for the OPG.

To achieve these targets, there were two fundamental tools: a rebate mechanism and a divesting schedule. The rebate mechanism mandated a price threshold (\$38/MWh), above which OPG must rebate most of its revenue back to consumers. The price cap of \$38 was selected to "ensure that consumers of electricity in Ontario do not pay a wholesale price for electricity much in excess of an average 3.8 cents per kilowatt hour, which together with a competition transition charge (CTC) or its equivalent, transmission, distribution, and other charges should amount to no more than current average retail electricity prices in the province of 7.2 cents per kilowatt hour" (MDC, 1998: Market Power, pages 1-4)<sup>15</sup>. The price of 3.8 cents per kilowatt hour was thought to be OPG's average production cost.

The rebate mechanism can be written as:

$$rebate = \left( \frac{\sum P_i Q_i^*}{\sum Q_i^*} - P_{cap} \right) * \sum Q_i^* \quad (2.1)$$

Where  $P_i$  is the market hourly price in hour  $i$ ,  $Q_i^*$  is the Contract Required Quantity (CRQ) for hour  $i$ , and  $P_{cap}$  is the price cap \$38/MWh. The Contract Required Quantity was derived through computer simulations supervised by the Market Design Committee. All suppliers were assumed to bid competitively except the OPG who was assumed to bid at average price \$38/MWh.

---

<sup>15</sup> Because the retail price prior to deregulation was regulated and may have been below the real average cost, the price cap may have been too low.

The rebate mechanism was also coupled with a credit clause called the Price Spike Adjustment (*PSA*) that protects the OPG from refunding the revenue that it hasn't earned. In particular, if the hourly market price is higher than \$125/MWh and at the same time the actual production of the OPG is less than *CRQ*, the OPG will not be required to rebate the following portion of its revenue:

$$PSA = (P_i - \$125)(Q_i^* - q_i) \quad (2.2)$$

where  $q_i$  is the actual production in hour  $i$  and must be less than the required quantity  $Q_i^*$ . The Adjustment clause raised concerns that the OPG could strategically make use of it by withholding its quantity and bid above \$125 in order to receive the credit.

The capacity decontrol schedule (which was abandoned in 2004) consisted of two stages, each with a mandatory divestiture schedule for each type of generation. The first stage was a period from market opening to its 42<sup>nd</sup> month, and the second stage was thereafter up to the 10th year after market opening. The two types of capacity include Tier 1, which consists of all hydroelectric and nuclear generation located in Ontario, and Tier 2, which includes all other generation in Ontario, the transmission capacity of the inter-ties, and demand-side bidding.

By the end of 42 months after the wholesale market opened, the OPG was obliged to divest the greater of:

1. enough of its generating output so that its share of the Tier 2 market is no greater than 35%;
2. 4,000 MW of Tier 2 capacity.

To motivate OPG to decontrol its capacity, every MW capacity sold was to result in a 1.1 MW reduction in the CRQ. However, to promote more diversified ownership, the capacity sold would not be counted either towards OPG's target or in reduction if the transferee had effective control over more than approximately 25% of either: 1) total in-service Tier 2 capacity, or 2) total in-service Tier 1 and Tier 2 capacity. By the end of ten years after open access began in Ontario, the OPG was to decontrol enough of its generating output to bring its share of the total Tier 1 and Tier 2 capacity to no more than 35%.

As stated before, the divesture provisions of the MPMA were not followed and were formally eliminated under the new provincial government electricity legislation in 2004. For this reason, the analysis will focus on the price cap and rebate mechanism embodied in the MPMA.

In a one-shot game, all market participants produce to the point at which their marginal revenue equals their marginal cost, regardless of the market structure. In a monopoly or oligopoly market, each firm faces a downward-sloped marginal revenue curve, which lies below the market demand curve. The market equilibrium price is higher than the competitive benchmark. In contrast, in a perfectly competitive market, the marginal revenue of each is equal to the market equilibrium price. Each firm will generate electricity until marginal cost equals that price, or maximum capacity is reached.

Viewed in terms of market shares, the Ontario wholesale electricity market resembles a dominant-firm-with-a-competitive-fringe model. The dominant firm, OPG, accounts for about 70% of the market, Bruce Power (nuclear units) for 21%, non-utility generators (or NUG's) for 6%, and independent producers for 3%. However, the market can be viewed as monopoly because Bruce Power is not price responsive (nuclear units can not quickly ramp up and down to respond to a price signal) and the non-utility generators have long term contracts with the provincial government at a fixed price, which is generally higher than the prevailing wholesale price, and are thus not price sensitive<sup>16</sup>.

There are many factors that could complicate a generator's bidding strategy. First, a firm's strategy relies on its position as a provider of default service or contract position (Fehr and Harbord, 1992, Wolak, 1999). A firm with a large contract commitment or large percentage of default load to serve compared to its generation capacity tends to bid more competitively, putting downward pressure on the spot price. It has been said that the OPG had about 30% of its production under contract at the time when the market was opened, but there is no such information in the public domain, and thus we cannot assess the strategic impact of those contracts. However, it is foreseeable that OPG's bidding strategy would be significantly affected by its contract position. Second, transmission capacity has a significant impact on offer strategies (Borenstein, Bushnell, and Knittle, 1999). In a spatial market, limited transmission capacity can provide generators with incentives to restrict their output in order to congest transmission into their area of

---

<sup>16</sup> Those non-utility generations may be responsive to (input) fuel price for two reasons. First, a NUG may produce more power when the fuel price is lower as long as it produces the contracted quantity. Second, it might sell back the fuel to the energy market if the fuel price is so high that selling back the fuel makes more money than consuming the fuel to generate power. For detailed implication in this case, see the MSP reports ([www.ieso.com](http://www.ieso.com))

dominance. With sufficiently large transmission capacity, oligopolists act more aggressively and transmission lines will actually seem overbuilt and underused. Third, market rules may affect firms' bidding behaviour (Wolak and Patrick, 1997, MSP, December 2005, June 2006). For example, generators in California were not allowed to sign long-term bilateral contracts in the initial stage of market opening, and thus were exposed to the volatile real time price. As a result, the generators had to bid higher to include a risk premium. In the Ontario case, the uniform price system induces intertie traders to target the congestion payment which may have no value to the market (MSP, December 2005, June 2006).

In a one-shot game, the average price is the single period market-clearing price, and  $CRQ$  is a single quantity. The rebate mechanism can thus be simplified as:

$$rebate = (P - P_{cap}) * Q^* \quad (2.3)$$

where  $P$  is the real time market price,  $P_{cap}$  is the price cap or price threshold, and  $Q^*$  is the  $CRQ$ . The rebate would flow back to consumers if  $P > P_{cap}$ . Otherwise, there would be no refund at all.

Hydroelectric generation, which accounts for one quarter of the total electricity supply in Ontario, is another interesting topic. The marginal cost of a hydroelectric unit is extremely low. Engineering estimates of the marginal cost of hydro electricity

production fall between \$2-5/MWh<sup>17</sup>. Adding some cost associated with environmental regulation, the total average cost would be around \$15/MWh. Given that the opportunity cost of hydroelectric units is irrelevant in a one shot game, we will not discuss it here, but will assess it in more detail in the next chapter.

## 2.4 The Model

The starting point of the static model is primarily based on an analysis conducted by the Market Assessment Unit (MAU) of the Independent Electricity System Operator (IESO) (MSP, 2002). The MAU is responsible for the daily monitoring activities of the Ontario wholesale electricity market under the guidance of the Market Surveillance Panel of the Ontario Energy Board.

In this model, there are two players in the market: a monopoly firm and a regulator. The firm is risk neutral and maximizes its profit only while the regulator maximizes social welfare defined as aggregate consumer and producer surplus. The monopolist has a quadratic cost function:  $C(q) = a + bq + \frac{c}{2}(q)^2$ , where  $q$  is electricity production, and  $a$ ,  $b$ , and  $c$  are constant coefficients. The cost function resembles the heat rate function of a fossil generator, which is widely used in the electricity industry. Therefore  $a$  represents the start-up cost of each commitment (in the current one shot game,  $a$  also represents the fixed cost of the station), and  $b$  and  $c$  stand for the incremental cost. According to the

---

<sup>17</sup> In a multi-period model, however, there is opportunity cost since the monopolist or any firm that owns hydro resources can optimize its hydro production across different periods so that it can maximize profit by shifting hydro-electric production from low price hours to high price hours.

technology of existing fossil units, a typical  $b$  is 6.0~7.5 times the fuel price per Giga Joule, and  $c$  is 0.0002~0.0003 times the fuel price.

The demand for electricity is assumed linear:  $P = A - Bq$ , where  $A$  and  $B$  are constant.

Thus, the monopsony power of buyers (Acton and Besen, 1987) is excluded. Market demand is assumed to include export demand. The possibility of importing is assumed to exist but only the monopolist can import. The monopolist's decision to import is considered later in this chapter.

If the price cap is so high that it is not binding, the profit function is of the unconstrained monopoly type:  $\pi = Pq - a - bq - \frac{c}{2}(q)^2$ . Profit maximization requires marginal revenue  $MR^u = \frac{d(Pq)}{dq} = A - 2Bq$  equal to marginal cost  $b + cq$ . The outcome will be the standard monopoly equilibrium (point  $A$  in Figure 2.1):  $q = \frac{A-b}{2B+c}$  and  $p = \frac{AB+Ac+Bb}{2B+c}$ .

If the price cap is binding, the monopolist's profit function becomes

$\pi = Pq - a - bq - \frac{c}{2}(q)^2 - (p - P_{cap})Q^*$ . As a result, marginal revenue is

$MR^c = A - 2Bq + BQ^*$ . If  $q < Q^*$ ,  $A - 2Bq + BQ^* > A - Bq$ , i.e.  $MR^c$  will be above

demand  $P(q) = A - Bq$ . If  $q \geq Q^*$ , however,  $MR^c$  will be below demand. Therefore, the upper part of the  $MR$  curve will cross the demand curve from above at  $Q^*$  (Figure 2.1). If the market price is lower than or equal to the price cap, however, the price cap will not be binding. The  $MR$  curve will remain the same as in the unrestricted monopoly case.

Therefore, the market power mitigation mechanism disconnects the  $MR$  curve

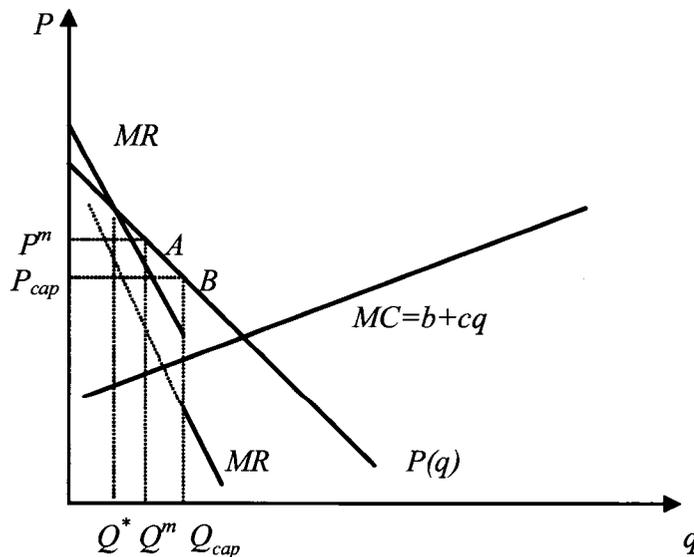
at  $Q_{cap} = \frac{A - P_{cap}}{B}$ : on the left side of  $Q_{cap}$ ,  $MR$  is  $A - 2BQ_{cap} + BQ^*$ , and on the right side,  $MR$  is  $A - 2BQ_{cap}$ .

Since the implicit goal of the MPMA is to induce the monopolist to bid at the price cap, the  $MC$  curve should intersect the  $MR$  curve at  $Q_{cap}$ . That is, marginal cost should be between the upper marginal revenue and the lower marginal revenue at  $Q_{cap}$ , i.e.

$A - 2BQ_{cap} \leq b + cQ_{cap} \leq A - 2BQ_{cap} + BQ^*$ , which can be rearranged as

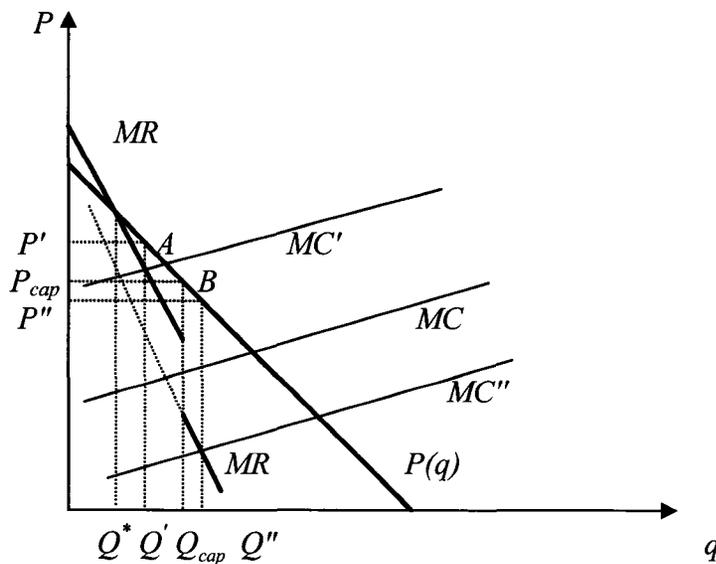
$\frac{1}{2}(A + b + cQ_{cap} - BQ^*) \leq P_{cap} \leq \frac{1}{2}(A + b + cQ_{cap})$ . The first inequality says that given a price cap, there are infinite numbers of  $Q^*$  that satisfy the condition as long as  $Q^*$  is sufficiently large. This leaves the regulator a lot of flexibility to maximize its own objective function.

**Figure 2.1: Equilibrium under the MPMA**



The rebate can shift the  $MR$  curve up and down by changing the combination of price cap and  $CRQ$ , and thus may make the  $MR$  intersect the  $MC$  at different quantities. For example, if marginal cost is high like  $MC'$  in Figure 2.2, the equilibrium price  $P'$  will be higher than the cap. While if the  $MC$  is low like  $MC''$ , the market equilibrium price  $P''$  will be lower than the cap, which reduces to a simple monopoly case. If the marginal cost curve is located like  $MC$ , the equilibrium price would be exactly the price cap.

**Figure 2.2: High or Low Cost Equilibria under the MPMA**



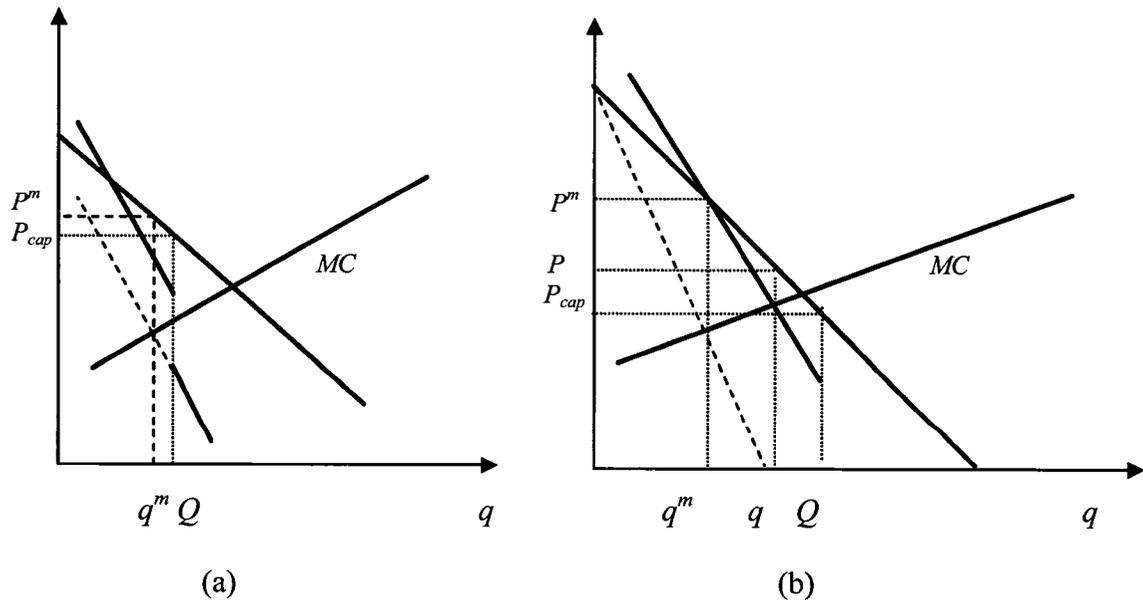
## 2.5 The Welfare Impact of the MPMA

If the price cap is higher than the pure monopoly price, for example, in the case with  $MC=MC''$  in Figure 2.2, the MPMA will have no impact both on the monopolist's

strategy and on social welfare. This would be the worst case with an implementation of nonbinding MPMA.

In contrast, with an effective MPMA, i.e. the price cap is lower than the monopoly price, there are two possible outcomes: one is the case of panel (a) in Figure 2.3, where the market equilibrium price is exactly equal to the price cap, and the other one the case of panel (b), where the market price is above the price cap. In the first case, the upper  $MR$  at  $Q$  must be greater than or equal to  $MC$ , i.e.  $A - 2BQ + BQ^* \geq b + cQ$ . Also, the lower  $MR$  at  $Q$  must be smaller than or equal to  $MC$ , i.e.  $A - 2BQ \leq b + cQ$ . Recall that under the unconstrained monopoly case (i.e. no MPMA in place),  $MR = MC$ , i.e.,  $A - 2Bq^m = b + cq^m$ , implying  $q^m \leq Q \leq \frac{Q^* + 2q^m}{2}$ . That is, output under MPMA is now greater than the monopoly output. Social welfare is improved.

**Figure 2.3: Welfare Improvement under the MPMA**



In the case of panel (b), the upper part of  $MR$  intersects the  $MC$  curve, i.e.

$A - 2Bq + BQ^* = b + cQ_{cap}$ , implying that  $q = q^m + \frac{1}{2}Q^*$ . Once again output is greater than the monopoly output; social welfare is improved.

Thus we have Proposition 1.

**Proposition 1** *The price cap mechanism in the MPMA is welfare-improving as long as it is lower than the unrestricted monopoly price and the  $CRQ$  is non-zero, regardless of the efficiency type of the monopolist.*

A simple  $MC$  may not reflect the whole picture of current cost structure of the OPG. In fact, after many years of no new investment, OPG's existing generation facilities became aged and the maintenance cost and other fixed costs have increased. As a result, OPG may now produce at a quantity at which the  $MC$  is greater than the historical average cost ( $AC$ ). Given that the pre-determined  $CRQ$  is 90% of the estimated quantity, OPG will produce when the market price equals its average cost, and the equilibrium price may not be the price cap.

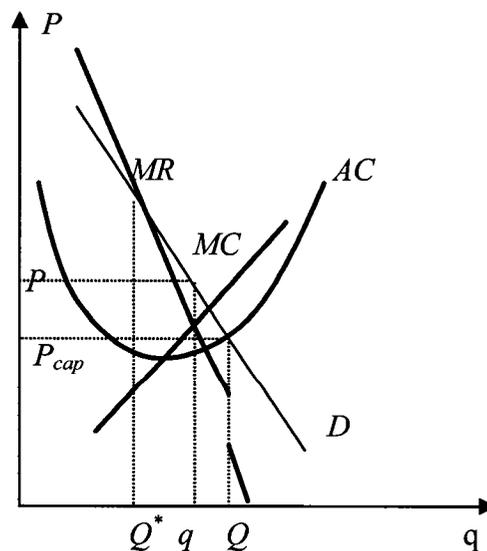
Under the MPMA, the price cap was set to approximate the average cost of electricity production in Ontario, i.e.,  $P_{cap} = AC = \frac{c(Q)}{Q}$ , where  $Q$  is the demand satisfied at which the firm breaks even. The  $CRQ$  is set to 90% of  $Q$ .

A necessary condition for the monopolist to bid the price cap is that the  $MC$  curve must intersect  $MR$  in the discontinuity, i.e. the marginal cost must be less than the upper  $MR$ , but greater than the lower  $MR$  at  $Q$  (Figure 2.1). In mathematical form,

$\frac{\partial P(Q)}{\partial q} Q + P(Q) \leq MC(Q) \leq 0.1 \frac{\partial P(Q)}{\partial q} Q + P(Q)$  because  $Q^* = 0.9Q$ . The left hand inequality can be ignored if an effective price cap is below the unconstrained monopoly price. Since  $\frac{\partial P(Q)}{\partial q} < 0$  and  $P(Q) = P_{cap} = \frac{C(Q)}{Q}$ , the necessary condition for bidding the price cap can be reduced to  $MC(Q) \leq P(Q) = \frac{C(Q)}{Q}$ . That is, marginal cost must be less than or equal to average cost, implying that the demand shouldn't be greater than the quantity at which the average cost is minimized.

If the quantity demanded at the cap lies to the right of the minimum point on the average cost curve, the monopolist will produce where the  $MR$  is equal to the  $MC$ , rather than where the demand curve crosses the average cost curve (Figure 2.4). In this case, the MPMA will fail to achieve the desired outcome, i.e. to induce the monopolist to bid at the price cap.

**Figure 2.4: Failure of the MPMA**



**Proposition 2**

*The MPMA is second best only if production is below the level at which average cost reaches its minimum. If output is such that average cost is above its minimum point, the monopolist's profit-maximizing bid is higher than the price cap.*

**2.6 Impact of the Price Spike Adjustment (PSA)**

In the MPMA, the *PSA* clause stipulates a threshold price of \$125/MWh, above which OPG's refund to consumers will be automatically reduced if its output is less than the *CRQ* at the same time. This clause was initially designed to protect the OPG from refunding revenue that it didn't earn, but leads to concerns from market participants and observers that the clause may induce the OPG to strategically withhold its capacity below the *CRQ* level.

With the *PSA* in place, the monopolist's revenue function becomes:

$$R = P(q)q - (P(q) - P_{cap})Q^* + (P(q) - 125)(Q^* - q) \quad (2.4)$$

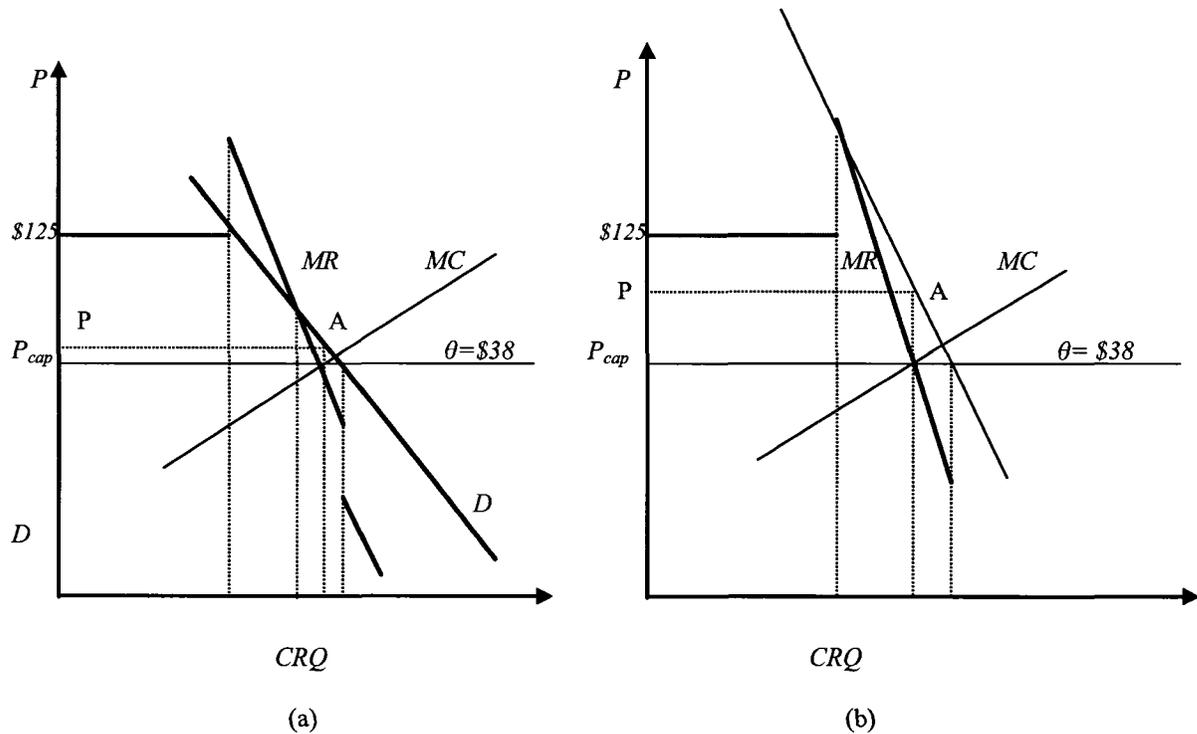
Substituting demand function  $P(q) = A - Bq$  and the price cap  $P_{cap} = 38$  into (2.4) yields:

$$R = 125q - 87Q^* \quad (2.5)$$

The marginal revenue  $\frac{dR}{dq}$  is thus a constant \$125. That means the *MR* curve is a horizontal line when the *PSA* applies. The effect of *PSA* is depicted in Figure 2.5. The *MR* curve for the whole production range is now divided into three parts: a horizontal line at \$125, a downward-sloping part and part of the original *MR* curve.

There are two possible situations on how the  $MR$  curve posits. One is when the  $CRQ$  is greater than the demand level at  $\$125/\text{MWh}$  (panel (a) in Figure 2.5). And another one is when the  $CRQ$  is less than the demand level at  $\$125/\text{MWh}$  (panel (b)).

**Figure 2.5: Effect of the Price Spike Adjustment**



Because a monopolist will always choose to produce at the point at which  $MR$  equals  $MC$ , it will have no incentive to bid above  $\$125$  as long as its marginal cost is lower than  $\$125$ . That is, the PSA will have no impact on the monopolist's strategy, and thus we have the following proposition.

**Proposition 3** *The Price Spike Adjustment (PSA) has no impact on the monopolist's bidding strategy*

## 2.7 Multiplicity of Optimal Policies

The socially optimal price is considered to be the price that maximizes the sum of consumer and producer surplus. In our model with linear functions, the maximization requires that the market equilibrium price equals marginal cost, i.e.  $P^o = b + cQ^*$ .

According to Propositions 1 and 2, the price cap must be set to be  $P_{cap} = P^o = \frac{Ac+Bb}{B+c}$ ,

and  $Q = Q^o = \frac{A-b}{B+c}$ <sup>18</sup>.

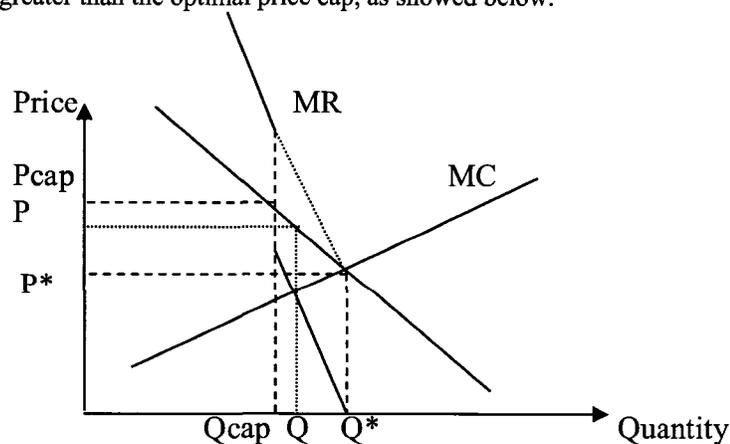
To induce the monopolist to bid at the price cap, it must be the case that the  $MC$  curve must intersect the  $MR$  at the disjoint point  $Q$ , i.e.  $A - 2BQ \leq b + cQ \leq A - 2BQ + BQ^*$ ,

Solving this inequality yields  $Q^* \geq Q^o$ , i.e.  $Q^* \geq \frac{A-b}{B+c}$ .

Therefore, the optimal price cap is equal to marginal cost, and the optimal  $CRQ$  must be greater than or equal to  $\frac{A-b}{B+c}$ . In particular, a higher price cap combined with a lower  $CRQ$

is needed for a high cost firm, while a lower price cap in conjunction with a higher  $CRQ$

<sup>18</sup> If the  $CRQ$  is set to be the optimal level but the price cap above the optimum price, the equilibrium price would be greater than the optimal price cap, as showed below.



should be imposed on a more efficient firm. In the case of high demand level or a low elasticity of demand, a high  $CRQ$  should be enforced. In contrast, with a low demand level or a high elasticity of demand, a low  $CRQ$  should be imposed.

The striking finding here is that the optimal  $CRQ$  must be greater than some threshold. The implication is that for the policy-maker who prefers a reasonably low and stable price, the best way to assure the target would be to set a relatively high  $CRQ$ , rather than a low one. Because there will be no rebate when the price is less than or equal to the price cap, a high  $CRQ$  won't impact the monopolist's profit if the market price is no greater than the cap. However, if the price is greater than the pre-specified cap, the high  $CRQ$  will serve as a punishment instrument.

If  $CRQ$  is optimally set to be  $\frac{A-b}{B+c}$ , the monopolist's profit function can be written as:

$$\pi = \begin{cases} -a + \frac{(A-b)^2}{2(2B+c)} & \text{if } P_{cap} > \frac{AB+Ac+Bb}{2B+c} \\ -a + \frac{(P_{cap}-b)(A-P_{cap})}{B} - \frac{c(A-P_{cap})}{2B^2} & \text{if } \frac{Ac+Bb}{B+c} < P_{cap} \leq \frac{AB+Ac+Bb}{2B+c} \\ -a + \frac{A-c}{B+c} [P_{cap} - b - \frac{c(A-b)}{2(B+c)}] & \text{if } P_{cap} \leq \frac{Ac+Bb}{B+c} \end{cases} \quad (2.6)$$

The profit function is positively related to the price cap. In particular, if the price cap is greater than the monopoly price, the price cap will not be binding and the firm will make the (unconstrained) monopoly profit. If the price cap is moderate, the firm can make a positive profit, which is an increasing function of price cap. While if the price cap is too low, the firm will experience a loss. The profit function is illustrated in solid line in Figure 2.6.

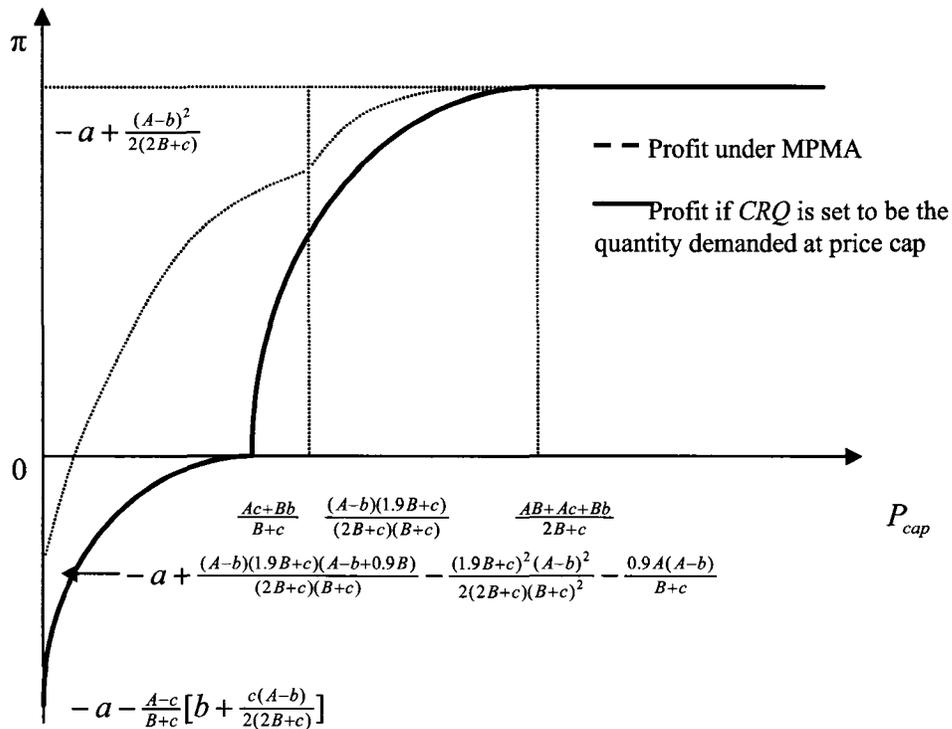
In contrast, if  $CRQ$  is set to be 90% of the socially optimal quantity  $\frac{A-b}{B+c}$ , as stipulated in the MPMA, the profit function will become

$$\pi = \begin{cases} -a + \frac{(A-b)^2}{2(2B+c)} & \text{if } P_{cap} > \frac{AB-Ac+Bb}{2B+c} \\ -a + \frac{(P_{cap}-b)(A-P_{cap})}{B} - \frac{c(A-P_{cap})}{2B^2} & \text{if } \frac{(A-b)(1.9B+c)}{(2B+c)(B+c)} < P_{cap} \leq \frac{AB-Ac+Bb}{2B+c} \\ -a + \frac{(A-b)(1.9B+c)(A-b+0.9B)}{(2B+c)(B+c)} - \frac{(1.9B+c)^2(A-b)^2}{2(2B+c)(B+c)^2} - \frac{0.9(A-P_{cap})(A-b)}{B+c} & \text{if } P_{cap} \leq \frac{(A-b)(1.9B+c)}{(2B+c)(B+c)} \end{cases} \quad (27)$$

Similar to the case with the  $CRQ$  being optimally chosen, the monopolist will bid at the monopoly price if the price cap is higher than the monopoly price. If the price cap is lower than  $\frac{(A-b)(1.9B+c)}{(2B+c)(B+c)}$ , the firm will make either negative or positive profit, depending on the magnitude of the price cap relative to demand and marginal cost and its fixed cost. If the price cap is in between, the profit is positive if the fixed cost is not too large and positively related to price cap. The dotted line in Figure 2.6 illustrates the profit under the 90% threshold.

As a result, requiring 90% of optimal quantity in the MPMA shifts up the profit curve. That being said, the price cap mechanism doesn't extract all the rent and doesn't lead to the socially optimal outcome either. To correct this inefficiency, in addition to the requirement that the optimal price cap be set at  $\frac{Ac+Bb}{B+c}$ , the optimal  $CRQ$  must be greater than or equal to  $\frac{A-b}{B+c}$ . Thus we have Proposition 4 as follows.

**Figure 2.6: Profit as a Function of the Price Cap**



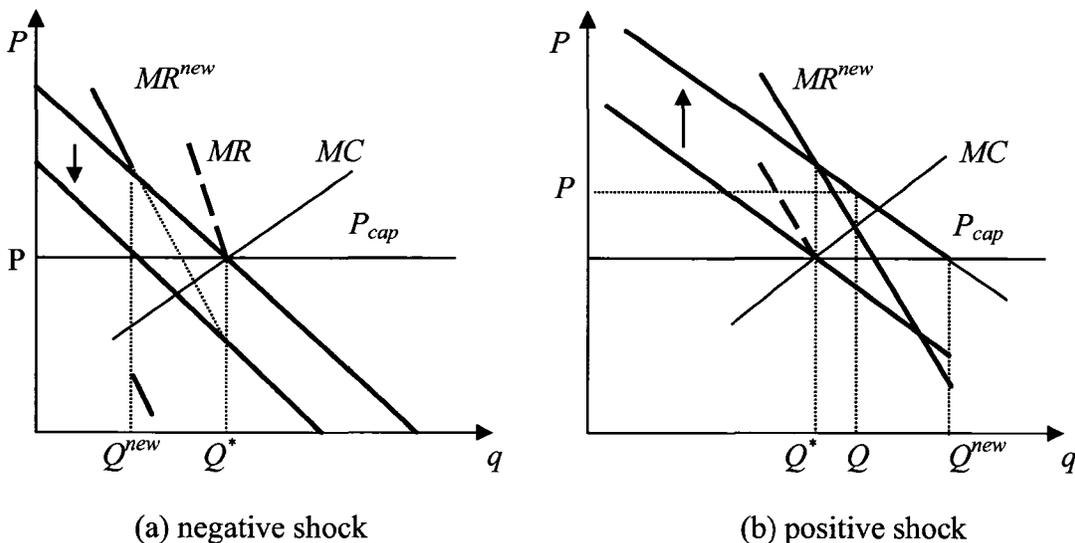
**Proposition 4 (multiplicity of optimal CRQ values)** As long as CRQ is greater than  $\frac{A-b}{B+c}$ , which is the first best quantity, social optimality can be achieved by simply setting the optimum price cap, implying that there are multiple CRQ's that could lead to the social optimum.

## 2.8 Demand Uncertainty

The “optimally” designed price cap scheme may be affected by a demand shock, for example, due to an unexpected change in weather conditions or an unexpected economic event. The *ex post* demand may be significantly different from the *ex ante* simulation or projection, and the monopolist would adjust its bidding strategy to respond to the change in demand.

First, let's look at the case where actual demand turns out to be less than expected. Call this a negative demand shock. The rebate mechanism is designed without considering uncertainty with  $P_{cap} = \frac{A_c + Bb}{B+c}$  and  $Q^* = \frac{A-b}{B+c}$  (for simplicity). When the demand shock ( $\varepsilon$ ) happens, the new demand function becomes  $P^{new} = A - \varepsilon - Bq$ , where  $P^{new}$  denotes the post-shock inverse demand function. Then the post-shock marginal revenue becomes  $A - 2Bq - \varepsilon$ , if  $q \geq Q^*$ , and  $A - 2Bq + BQ^* - \varepsilon$ , if  $q < Q^*$ . That is, the  $MR$  curve is shifted down by  $\varepsilon$ . Remember that the disjoint point on the  $MR$  curve corresponds to the quantity at which the price cap intersects the original demand curve. Thus, the disjoint point of the  $MR$  curve will also shift to the left. The corresponding quantity will be the new equilibrium quantity (Figure 2.7 (a)). At equilibrium, both the firm and consumers are worse off and the equilibrium leads to a deadweight loss.

**Figure 2.7: Demand Shocks and the Optimum**



In contrast, if the demand shock is positive, both parts of  $MR$  will shift up by  $\varepsilon$ . The

discontinuous point is now  $Q^{new}$  (Figure 2.7 (b)). Given  $MC$  unchanged, the upper  $MR$  curve will now intersect  $MC$ . As a result, the new equilibrium will be  $(P, Q)$ , which is higher than the optimum. In this case, it is obvious that the monopolist is making positive profit, but some consumers are rationed (although efficiently).

***Proposition 5 (effect of a demand shock)** If the contract is designed without taking into account demand uncertainty, the monopolist will be worse off under a negative demand shock while better off under a positive demand shock. Consumer surplus is never maximized, that is, there are some consumers whose consumption will be rationed.*

As shown above, an “optimal” policy under the certainty case won’t be optimal in the presence of demand uncertainty. There is plenty of research on optimal regulation under demand uncertainty (e.g. Lewis and Sappington, 1988a, 1988b, Keenan and Keenan, 1984, Saving and De Vany, 1981, Biglaiser and Ma, 1995, Wu, 1979). Most studies make use of a multi-period framework, and assume the regulated firm has better information about demand. Lewis and Sappington (1988a) find that as long as marginal cost is non-decreasing in quantity, marginal cost pricing would be the first best and feasible. However, when marginal cost declines with quantity, the risk neutral regulator is unable to derive any benefit from the firm’s superior knowledge, and a single price is established that is invariant to demand. In contrast, Keenan and Keenan (1984) assume the regulator is risk averse and find that marginal cost pricing fails to remain the first best. They further demonstrate that with demand independence, the first best policy is a

limiting case of second best policy and also takes the form of a simple inverse elasticity rule.

In the Ontario case, the assumption of asymmetric information on demand between the regulator and the firm is an implausible assumption since both the regulator and the firm actually have essentially identical techniques to forecast the electricity demand (due to deregulation of a previously vertically integrated and regulated monopoly) and can observe demand *ex post* (-- demand for electricity must be met all the time and was published every five minutes by the system operator, IESO). Thus in the analysis that follows, we will assume that both the regulator and the firm have the same expectations about demand and negotiate a contract in the first period, and the firm decides how much to produce before actual demand is realized. The market price is determined in the second period by the realized demand function. That is, the model is a simplified type of Lewis and Sappington regulation (1988a), but we extend their model to a second best case.

Assume there are only two possible levels of demand: low demand  $P_l = A - \varepsilon - Bq$  with probability  $\mu$  with  $\mu > 0$ , and high demand  $P_h = A + \varepsilon - Bq$  with probability  $1 - \mu$ . The cost function is assumed to be public knowledge, and demand can be verified *ex post*. As a result, the optimal policy would be *ex post* marginal cost pricing, as shown in Lewis and Sappington<sup>19</sup>.

---

<sup>19</sup> In our two-period model where the hydro resource is included, however, a marginal cost pricing of fossil unit won't be the first best.

When a rebate mechanism is in place rather than the requirement of marginal cost pricing, however, there is no optimal rebate policy. Assume there is no income effect. The regulator's problem can be written as (both rebate payment and subsidy/tax cancel out each other):

$$\begin{aligned} \text{Max } : W = & \mu \left[ \int_0^q (A - \varepsilon - Bq - b - cq) dq - a \right] + \\ & (1 - \mu) \left[ \int_0^q (A + \varepsilon - Bq - b - cq) dq - a \right] \end{aligned} \quad (2.8)$$

Thus the optimal rebate mechanism is  $P_{cap}^* = \frac{Ac+Bb+c(1-2\mu)\varepsilon}{B+c}$  and  $CRQ^* = \frac{A-b+(1-2\mu)\varepsilon}{B+c}$ . If there exists constant marginal cost, i.e.  $c=0$ , then  $P_{cap}^* = b$  and  $CRQ^* = \frac{A-b+(1-2\mu)\varepsilon}{B}$ . To allow the firm to break even, the optimal transfer/tax should be  $T^* = \frac{c}{2}(CRQ^*)^2 - a$  (a transfer to the firm when  $T^*$  is positive, or a lump-sum tax from the firm when  $T^*$  is negative). There is no surprise here: this optimal mechanism is determined by the expected demand. Specifically, when a high demand is more likely to emerge (a small  $\mu$ ), a high price cap and high  $CRQ$  would be required. In contrast, when a low demand is more likely to occur (a large  $\mu$ ), a low price cap and low  $CRQ$  are needed.

After having signed the contract based on the expectation of demand, the monopolist can now choose its output level and price for the coming period. Note that the firm has to decide an output level before demand is realized, while in Proposition 5 demand is perfectly foreseen by the firm, but not by the regulator. It is apparent that the price cap is between  $\frac{A-\varepsilon-b}{B+c}$ , the competitive price at low demand, and  $\frac{A+\varepsilon-b}{B+c}$ , the competitive price at high demand, and the planned output between the two corresponding output levels. Thus

the equilibrium price will be less than the low demand price, while greater than the high demand price, as shown before. Consequently, the firm's profit problem is:

$$\begin{aligned} \text{Max : } \pi = & \mu[(A - \varepsilon - Bq)q - a - bq - \frac{c}{2}q^2 + T^*] \\ & + (1 - \mu)[(A + \varepsilon - Bq)q - a - bq - \frac{c}{2}q^2 - (A + \varepsilon - Bq - P_{cap}^*)CRQ^* + T^*] \end{aligned} \quad (2.9)$$

The privately optimal output level is thus:

$$q_f^* = (1 - \frac{B\mu}{2B+c}) * CRQ^* \quad (2.10)$$

which is less than the socially optimal  $CRQ^*$ , as long as  $\mu$  is positive. Consequently, the equilibrium price is greater than the socially optimal price.

The implication is that because the monopolist faces a positive possibility of a low demand, it will suffer a loss in such a situation. By producing less than the CRQ, it lowers the risk of a large loss when demand turns out to be low although it will make less profit when demand turns out to be high.

Therefore we have following proposition.

***Proposition 6 (Optimal policy under demand uncertainty)***

1. *The optimal policy is to require the monopolist to price at marginal cost, and calculate the transfer or tax ex post.*
2. *There exists no optimal rebate mechanism under demand uncertainty. In fact, the monopolist will produce less than the socially optimal level.*

## 2.9 Second Best Rebate Mechanism Under Demand Uncertainty

A first best regulation may not be achievable due mainly to political constraints. In fact, regulation is usually second or third best. That is, a regulator tends to maximize consumer's welfare subject to the requirement that the regulated firm could break even. Second best regulation is most important when there exists economies of scale, i.e. a firm cannot survive if first best pricing is required and no transfer is allowed.

In the current case, assume the fixed cost portion,  $a$ , is large enough so that the cost function exhibits economies of scale. Then the second best model can be written as:

$$\begin{aligned} \text{Max } W &= \mu \int (A - \varepsilon - Bq) dq + (1 - \mu) \int (A + \varepsilon - Bq) dq & (2.11) \\ \text{st. : } & \mu [A - \varepsilon - Bq]q - a - bq - \frac{c}{2}q^2 + (1 - \mu) [(A + \varepsilon - Bq)q - a - bq - \frac{c}{2}q^2] \geq 0 \end{aligned}$$

It is straightforward to derive the break-even output  $q^*$  and price  $P^*$ .

$$q^* = \frac{A - b + (1 - 2\mu)\varepsilon + \sqrt{[A + (1 - 2\mu)\varepsilon - b]^2 - a(2B + c)^2}}{2B + c} \quad (2.12)$$

$$P^* = \frac{(B + c)[A + (1 - 2\mu)\varepsilon] + Bb - B\sqrt{[A + (1 - 2\mu)\varepsilon - b]^2 - a(2B + c)^2}}{2B + c} \quad (2.13)$$

If  $a=0$  and  $c=0$ , i.e. there exist constant returns to scale, then the second best is the first best  $P^* = b$  and  $q^* = \frac{A - b + (1 - 2\mu)\varepsilon}{B}$ .

Given any price cap and  $CRQ$ , the optimal strategy for the monopoly is Problem (2.9) and thus the optimal output is (2.10). To induce the firm to produce at the second best

output level,  $q_f^*$ , the optimal output level from Problem (2.9), should be equal to  $q^*$ .

That is,

$$\left(1 - \frac{B\mu}{2B+c}\right) * CRQ^* = \frac{A+(1-2\mu)\varepsilon-b}{2B+c} + \sqrt{\left(\frac{A+(1-2\mu)\varepsilon-b}{2B+c}\right)^2 - a}$$

The second best CRQ, denoted as  $CRQ_s^*$ , is

$$CRQ_s^* = \frac{A-b+(1-2\mu)\varepsilon + \sqrt{[A-b+(1-2\mu)\varepsilon]^2 - a(2B+c)^2}}{2B+c-B\mu} \quad (2.14)$$

which can be further written as  $\frac{2B+c}{2B+c-B\mu} q^*$ . It is apparent that  $CRQ_s^* > q^*$  given a positive  $\mu$ . If the possibility of a low demand is very small, the second best CRQ converges to the break even output.

The second best price cap is

$$P_s^* = \frac{(B+c)[A+(1-2\mu)\varepsilon] + Bb - AB\mu - B\mu(1-2\mu) - B\sqrt{[A+(1-2\mu)\varepsilon-b]^2 - a(2B+c)^2}}{2B+c-B\mu} \quad (2.15)$$

which is obviously less than  $p^*$  because  $CRQ_s^* > q^*$ .

Alternatively, as illustrated in the first best solution, second best regulation can be achieved by setting the second best price as a price ceiling. Given the tendency for the firm to price above the second best price level, the fixed price cap can serve as the second best regulation instrument.

However, a flexible price cap  $P_s^*$  may be superior to a fixed price under demand uncertainty. The reason is that when demand turns out to be high, the firm does not benefit from it because the price is fixed at the expected break even point, while when demand is low, the firm makes a loss. In contrast, a flexible price cap allows the firm some profit when demand is high and the firm bears all risks when demand turns out to be low. Furthermore, a flexible price cap is more compatible with a competitive market.

In summary, Proposition 7 follows.

**Proposition 7** (*second best regulation under demand uncertainty*)

1. *The second best CRQ is greater than the expected break even output. If the probability of a low demand is very small, the second best CRQ converges to the break even output. Correspondingly, the second best price cap is lower than the break even price.*
2. *An alternative to the rebate mechanism is to set the break even price as a fixed price cap, above which no offer is allowed.*
3. *The soft price cap and rebate mechanism is superior to a fixed price cap because the rebate mechanism allows profit sharing while the fixed price puts all risks on the firm.*

## 2.10 Supply Uncertainty 1: Fuel Cost

In the electric power generation industry, the main source of supply uncertainty is fuel cost<sup>20</sup>. Most electricity generators are using one of uranium, coal, natural gas or water as their primary fuel. Except for nuclear generators, which use uranium and may have very long term contracts of supply with a fixed price, generators using other fuel types are more likely exposed to volatile supply and prices. For example, coal and natural gas prices (especially natural gas) may fluctuate day by day. The generation capacities of hydroelectric units are subject to the available water stock which depends on the weather conditions.

Assume the optimal policy is that  $P_{cap}^o = \frac{Ac+Bb}{B+c}$  and  $Q^o = \frac{A-b}{B+c}$ . Denote the new cost function as  $C(Q) = a + b_{new}q + \frac{c}{2}q^2$ , assuming that the change in fuel cost shifts the marginal cost up or down. The monopolist's objective function is thus:

$$\begin{aligned} \text{Max}_q : \quad \pi &= (A - Bq)q - a - b_{new}q - \frac{c}{2}q^2 - \text{rebate} & (2.16) \\ \text{where} \quad \text{rebate} &= \begin{cases} 0 & \text{if } A - Bq \leq P_{cap}^o \\ (A - Bq - P_{cap}^o) * Q^o & \text{if } A - Bq > P_{cap}^o \end{cases} \end{aligned}$$

If  $b_{new} < b$ , implying that the firm experiences an unexpected decrease in production cost, then it is straightforward that the equilibrium yields  $P = P_{cap}^o$  and  $q = Q^o$  due to the vertical discontinuity of the MR curve. Therefore the firm is still producing at the pre-

---

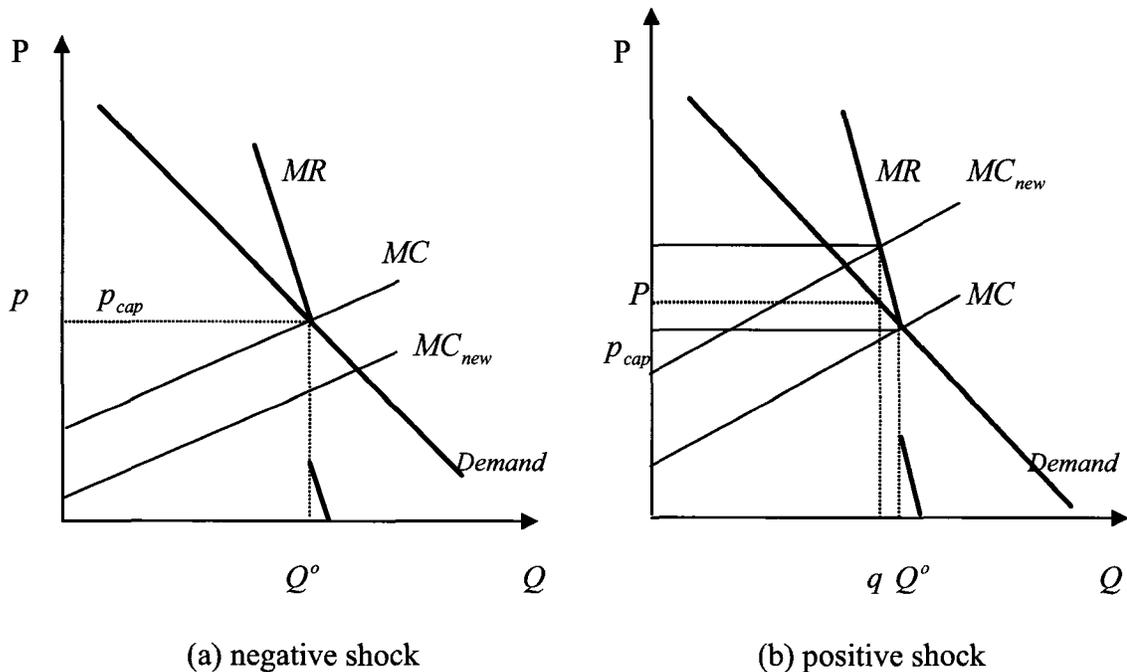
<sup>20</sup> There are also other supply shocks, such as unit outage, labor expense increase, tightened environmental regulation, etc.

shock optimum, but makes positive profit. Obviously, the outcome is not socially optimal: the generator is producing too little.

In contrast, when  $b_{new} > b$ , implying that the firm experiences an unexpected increase in fuel cost, the equilibrium output will be  $q = \frac{1}{2B+c}(A - b_{new} + B * Q^o)$  (a result of a binding price cap), which is less than the  $CRQ$ ,  $Q^o$ , (due to  $b_{new} > b$ ), but greater than the new socially optimal output  $q^o = \frac{A - b_{new}}{B+c}$ . As a result, the equilibrium price will be greater than the price cap, but lower than the marginal cost of producing the last unit of electricity. Substituting the equilibrium result into the profit function, one can see that the monopolist is now making negative profit. The outcome is not socially optimal either: the generator is producing too much.

These results are illustrated in Figure 2.8. In panel (a), a negative supply shock lowers the firm's marginal cost, leaving the monopolist a positive profit equivalent to the area between the two MC curves and between y axis and  $Q^o$ . In panel (b), the firm experiences a positive supply shock. The new equilibrium will be at  $(p, q)$ , where the market price is higher than the price cap, and the output is lower than the  $CRQ$ . Note that in both cases the equilibrium is not the social optimum. When only cost uncertainty exists and demand is observable, the optimal policy is marginal cost pricing (Lewis and Sappington, 1988a).

**Figure 2.8: Equilibrium with a Supply Shock**



**Proposition 8** *If the policy is designed without taking into account supply uncertainty, the equilibrium won't be socially optimal and the price cap will serve as the lower bound of market prices. If the supply shock is positive, e.g. an increase in fuel cost, the monopolist prices above the cap, experiences a loss and produces too much, while if the supply shock is negative, the monopolist prices at the cap, makes a profit and produces too little.*

## 2.11 Supply Uncertainty 2: Forced Outages

An electricity generator may experience sudden technical difficulties from time to time, i.e. forced outages, which makes the generator unable to produce electricity as planned.

To deal with a forced outage in real time, activation of operating reserve (capacity which is on line but not producing any power) or emergency imports by the system operator are required. That is, at the time of the outage, the generator cannot immediately respond to the outage, but can import to mitigate the price impact in later hours or periods – no immediate import strategy can be realized. Because a technical problem leading to an outage usually takes several hours to fix, the generator may purchase external electricity to fulfill its commitment if needed or maximize its profit as appropriate. The interaction of energy and operating reserve is very complicated and we will discuss the issue in Chapter 4; here we only discuss the impact of imports by the generator itself in a simple way.

Assume the import price is fixed (i.e. the external markets are so large that transactions between Ontario and those markets have no impact on external prices) and higher than the domestic price cap,  $P_{cap}$ , representing the fact that the real time prices in adjacent neighbours such as in New York are persistently higher than the Ontario price<sup>21</sup>. The domestic monopoly generator is assumed to have exclusive rights to import to or export from Ontario. Again, assume that the regulatory policy is *ex ante* optimal, i.e. the policy is designed to induce the monopolist to produce at the expected optimal level.

---

<sup>21</sup> The lower price in Ontario may reflect the fact that Ontario has more low cost generation such as hydro and coal fired generations. In contrast, in New York, the majority of plants are natural gas fired. Due to trade impediments and other constraints, the price in Ontario is generally different from the price in New York although they have converged (MSP, December 2005).

Denote  $\bar{q}$  as the maximum available capacity after an outage. If the firm simply produces at  $\bar{q}$ , the market price will be obviously greater than the cap and there is a rebate. The monopolist's profit is now:

$$\pi(\bar{q}) = (A - B\bar{q})\bar{q} - C(\bar{q}) - (A - B\bar{q} - P_{cap})CRQ \quad (2.17)$$

In contrast, the firm may instead import electric power to meet the shortage. In this case, marginal revenue in the home market must be equal to the purchasing price in the foreign market, i.e.  $A - 2B(\bar{q} + q_{imp}) + B * CRQ = P_{imp}$ , where  $q_{imp}$  is the imported quantity and  $P_{imp}$  is the import price. Rearranging the equation yields the import volume  $q_{imp} = (A - P_{imp} + B * CRQ) / (2B) - \bar{q}$ . Thus the monopolist's profit with imports is:

$$\begin{aligned} \pi(\bar{q} + q_{imp}) = & (A - B(\bar{q} + q_{imp}))(\bar{q} + q_{imp}) - C(\bar{q}) - P_{imp} * q_{imp} \\ & - (A - B(\bar{q} + q_{imp}) - P_{cap}) * CRQ \end{aligned} \quad (2.18)$$

Equation (2.18) will be greater than (2.17), if  $P_{imp} \leq A + B * CRQ - 2B\bar{q}$ , which is the marginal revenue at  $\bar{q}$ . That is, if the forced outage is large enough that marginal revenue at the maximum available capacity is greater than the import price, the monopolist will import rather than passively produce at the maximum capacity  $\bar{q}$ .

However, if the price is too high and the capacity is less than the  $CRQ$ , the Price Spike Adjustment (PSA) may apply. Under such circumstances, the monopolist doesn't have to

pay back the revenue earned from the high price. Thus the monopolist may have no incentive to import. Then the monopolist's profit is:

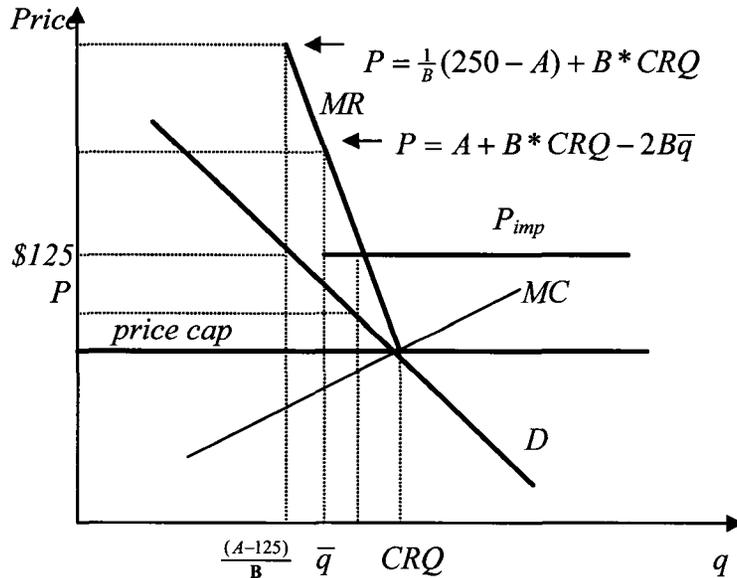
$$\pi(\bar{q}) = (A - B\bar{q})\bar{q} - C(\bar{q}) - (A - B\bar{q} - P_{cap}) * CRQ + (P(\bar{q}) - 125)(CRQ - \bar{q}) \quad (2.19)$$

As a result, equation (2.19) may be smaller than (2.18) if  $P_{imp} \leq \frac{1}{B}(250 - A) + B * CRQ$ , which is the marginal revenue at price \$125/MWh. That is, the firm has an incentive to import when the external price is sufficiently low. The result is illustrated in Figure 2.9. As long as the import price is lower than  $A + B * CRQ - 2B\bar{q}$ , or  $\frac{1}{B}(250 - A) + B * CRQ$ , whichever is smaller, the monopolist will import electricity. In this case, the domestic price  $\frac{1}{2}(A + P_{imp} - B * CRQ)$  (the market price at  $q = \bar{q} + q_{imp}$ ) would be lower than the import price, implying that the monopolist is losing money on the imports. At equilibrium, society is consuming more than the social optimum.

The monopolist will suffer losses in the outage situation, whether or not it purchases electricity from external jurisdictions. The implication is that importing in an outage situation may reduce the monopolist's loss, but will never eliminate it.

**Proposition 8** *If an outage reduces capacity below the CRQ, the monopolist may be better off by importing electricity to mitigate the price impact (eventually lower the rebate to consumers), even if the import price is higher than the domestic price. However it would not import if the external price is higher than  $\frac{1}{B}(250 - A) + B * CRQ$  or  $A + B * CRQ - 2B\bar{q}$ , whichever is smaller, either because the price is too high or because the Price Spike Adjustment insulates its revenue from the impact of a high price. The firm will suffer losses whether or not it imports electricity.*

**Figure 2.9: Imports with a Forced Outage**



## 2.12 Conclusions

There are many approaches to mitigating the market power of dominant firms.

Traditional regulation includes two extremes--cost plus and fixed price, and other incentive regulations in between. In most cases, regulation is subject to periodic reviews by the regulatory authorities, and the market price is rigid between reviews.

The traditional approaches lack flexibility and thus may have a significant adverse effect on the firm under regulation when the firm in question is facing a volatile market. For example, in a deregulated electricity market, all firms are bidding into a market that is cleared every five minutes, and the market price is extremely volatile due to the rapidly changing demand and supply conditions. A firm that is subject to a rigid regulation may

become unviable when the market turns out to be unfavourable, or make supra-normal profits when the market turns out to be favourable.

The rebate mechanism in the MPMA has some extent of built-in flexibility: the firm can bid above the price cap if needed, the *CRQ* is only 90% of the total output, and there are other protection clauses to ensure revenue sufficiency. All these designs are consistent with competition in a volatile market.

In summary, the theoretical analysis in this chapter shows that the price cap and rebate mechanism in the MPMA is welfare improving, and has a significant impact on a monopoly generator's bidding strategy. First, the rebate mechanism induces the firm to increase output beyond the unconstrained monopoly output, and thus increases market efficiency. Second, it requires payback of part of its extra revenue to consumers, which limits its incentive to bid too high. Third, under demand uncertainty, the second best price cap is lower than the expected break even price and *CRQ* is greater than the expected break even output. The rebate mechanism allows revenue sharing, leading to the conclusion that the MPMA is superior to a fixed price cap in this setting. Fourth, during a supply shortage period, the monopolist may purchase more expensive imports and sell at a lower price into the domestic market. Finally, there are multiple policies that yield the social optimum when there is no demand uncertainty. However, in the presence of demand uncertainty, there is no optimal MPMA. In this case, marginal cost pricing is the only optimal solution.

## References

1. Acton, J.P. and S. M. Besen (1987) "Assessing the Effects of Bulk Power Rate Regulation: Results from a Market Experiment," *Applied Economics* 19: 663-685.
2. Averch, H. and L. L. Johnson (1962) "Behavior of the Firm Under Regulatory Constraint", *American Economic Review* 52: 1053-1069.
3. Baron, D. and R. Myerson (1982) "Regulating a Monopolist with Unknown Costs", *Econometrica* 50: 911-930.
4. Biglaiser, G. and C.A. Ma (1995) "Regulating a Dominant Firm: Unknown Demand and Industry Structure", *Rand Journal of Economics* 26(1): 1-19.
5. Borenstein, S. J. Bushnell and C. Knittel (1999) "Market Power in Electricity Markets: Beyond Concentration Measures", *The Energy Journal* 20(4): 65-88.
6. Braeutigam, R. and J. Panzar (1989) "Diversification Incentives under 'Price-based and 'Cost-based' Regulation", *Rand Journal of Economics* 20: 373-391.
7. Fehr, and Harbord, (1992) "Long-Term Contracts and Imperfect Competitive Spot Markets: A Study of U.K. Electricity Industry", Oslo University, Department of Economics, Memorandum 14/1992.
8. Isaac, R.M. (1991) "Price Cap Regulation: A Case Study of Some Pitfalls of Implementation", *Journal of Regulatory Economics* 3:193-210.
9. Joskow, P. (1974) "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation", *Journal of Law and Economics* 17:291-327.
10. Joskow, P. and R. Schmalensee (1986) "Incentive Regulation for Electric Utilities", *Yale Journal on Regulation* 4:1-49.
11. Keenan, D.C. and T.K. Keenan (1984) "First and Second-Best Pricing with Unknown Demand", *Southern Economic Journal* 51(1): 80-88.
12. Laffont, J-J. and J. Tirole (1986) "Using Cost Observations to Regulate Firms" *Journal of Political Economy* 94: 614-641.
13. Laffont, J-J. and J. Tirole (1994) *A Theory of Incentives in Procurement and Regulation*, The MIT Press, Cambridge, Massachusetts.
14. Lewis, T.R. and D. Sappington (1988a) "Regulating a Monopolist with Unknown Demand", *American Economic Review* 78(5): 986-998.

15. Lewis, T.R. and D. Sappington (1988b) "Regulating a Monopolist with Unknown Demand and Cost Functions", *Rand Journal of Economics* 19(3): 438-457.
16. Market Design Committee (1998), *Third Interim Report to Minister of Energy, Science and Technology of Ontario*. Ministry of Energy, Government of Ontario
17. Market Surveillance Panel, *Market Monitoring Report*, December 2005, June 2006.
18. Market Surveillance Panel (2002) "Issue note: The Effect of The Rebate Mechanism on OPGI's Pricing Incentives" (IESO, Toronto).
19. Sappington, D. E. (2001) "Price Regulation and Incentives", in Martin Cave, Sumit Majumdar, and Ingo Vogelsang eds., *The Handbook of Telecommunications Economics* 1:227-296, North-Holland Publishing, Amsterdam
20. Saving, T.R. and A.S. De Vany (1981) "Uncertain Market, Reliability and Peak-Load Pricing", *Southern Economic Journal* 47(4): 908-923.
21. U.S. Department of Commerce (1991) *The NTIA Infrastructure Report: Telecommunications in the Age of Information*. Washington, D.C.: NTIA Special Publication.
22. Vogelsang, I. (2002) "Incentive Regulation and Competition in Public Utility Markets: A 20-year Perspective", *J. of Regulatory Economics* 22: 5-27.
23. Wolak, F. (1999) "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market", mimeo, Stanford University, Available at [www.stanford.edu/~wolak](http://www.stanford.edu/~wolak).
24. Wolak, F., and R.H. Patrick (1997) "The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market", POWER Working Paper PWP-047, University of California Energy Institute, Available at [www.ucei.berkeley.edu/ucei/PDFDown.html](http://www.ucei.berkeley.edu/ucei/PDFDown.html).
25. Wu, S.Y. (1979) "An Essay on Monopoly Power and Stable Price Policy", *American Economic Review* 69(1): 60-72.

## Chapter 3 A Two-period Model

In this chapter, we will introduce a two period model to investigate the strategic behaviour of a monopolist with energy-limited hydroelectric generation. The hydro resource is assumed to have no physical cost but is able to shift production between periods. This raises the possibility that the monopoly might strategically allocate the water resource for its own benefit.

An energy limited resource could also include fossil fuelled generation with emission limits or other environmental regulations (MSP, December 2005). The emission limits can be a yearly maximum emission of  $SO_2$ ,  $NO_2$ ,  $CO_2$ , and others. Other environmental regulation could include a maximum temperature of downstream water that goes through a fossil fuelled generation station. For example, a generator may use river or lake water a coolant and its effluent may increase the local water temperature. To protect the ecosystem of the lake or river, environmental authorities usually set a maximum temperature for lake or downstream river water and this may require the generator to reduce its use of water for coolant. In this analysis, however, we will confine our attention to energy-limited hydroelectric generation. The implications for other limitations on generation are similar.

Hydroelectric power accounts for about one quarter of the electricity supply in Ontario. OPG, the dominant generator, has a portfolio that includes both hydro and thermal generation. As explained in Chapter 1, the market power of OPG was constrained over

the period 2002 – 2005 by the Market Power Mitigation Agreement (MPMA). The analysis in this chapter is intended to predict the effect the MPMA would have on a hypothetical monopolist's allocation of its limited hydro generation resources between peak and off-peak periods.

Demand shifting is also an interesting topic since empirical analyses show that consumers (especially large consumers) do show significant consumption shifting away from high price hours (Heberlein and Warriner, 1983, Aigner and Hirschberg, 1985, Goldman et al, 2001). For example, large industrial consumers can significantly move their consumption from peak to off-peak (MSP, June 2003). However, the majority of consumers do not respond to price signals (Littlechild, 2001), for lack of real time meters, or because the price that they pay is regulated by the government, or because they don't have the equipment or skills to shift their consumption. Consumption substitution has become an increasingly interesting topic because the Ontario government is going to install "smart" meters (i.e. real time meters) for all residential consumers in the coming years, and a time variant price scheme will apply after they are installed. At that time, demand shifting may play a very important role in price determination. While demand side substitutability could be modelled by incorporating the second period price into first period demand, and vice versa, the model would be more difficult to solve but the basic results remain same.

### 3.1 The Model

There is relatively little published work on the exercise of market power by hydraulic generators (Ambec and Doucet, 2003, Bushnell, 2000). Bushnell (2000) models the allocation of hydro resources in a multi-period setting and derives an optimal allocation via computation simulation. In this chapter, we confine ourselves to a two period model, which allows us to derive many interesting insights into the regulation of hydroelectric generators in general and the application of the MPMA to hydroelectric generation in particular.

A monopoly generator is assumed to have two separate plants: one hydroelectric and one fossil-fuelled. The hydroelectric unit has a zero physical cost<sup>22</sup>, but is subject to a maximum available water stock. In fact, there may be many hydroelectric units/stations, as long as the total available water is limited. The assumption of a single station does not affect the general conclusions. The production of hydroelectricity is associated with an opportunity cost, which is the incremental cost of using fossil fuelled plants to serve its consumers. The fossil-fuelled plant is assumed to have a quadratic cost function:

$C(q^T) = a + bq^T + \frac{c}{2}(q^T)^2$ , as before. All parameters are constant and positive.

The rebate mechanism can be expressed in the following way:

$$rebate = \left( \frac{\sum (P_i * Q_i^*)}{\sum Q_i} - P_{cap} \right) * CRQ \quad \text{for } i = 1, 2 \quad (3.1)$$

---

<sup>22</sup> The cost that the OPG has to pay for the use of hydro resource is believed to be very low, but there is no specific information in the public domain.

where  $P_i$  is the real time market price in period  $i$ ,  $Q_i^*$  is the CRQ for period  $i$ , and  $P_{cap}$  is the predetermined price cap.

Therefore the monopolist's optimization problem can be written as follows:

$$\begin{aligned}
 \text{Max: } & \sum P_i * (q_i^H + q_i^T) - C_i(q_i^T) \} - \text{rebate} & (3.2) \\
 \text{s.t.: } & (1) \quad \sum q_i^H = \bar{q}^H \quad i=1,2 \\
 & (2) \quad P_i = A_i - B * (q_i^H + q_i^T) \\
 & (3) \quad C_i(q_i^T) = a + b * q_i^T + \frac{1}{2} c (q_i^T)^2 \\
 & (4) \quad \text{rebate} = \begin{cases} ((\sum P_i * Q_i^*) \cdot \sum Q_i^8 - P_{cap}) * \sum Q_i^* & \text{if } \sum P_i * Q_i^* / \sum Q_i^* > P_{cap} \\ 0, & \text{otherwise} \end{cases}
 \end{aligned}$$

where  $q_i^H$  and  $q_i^T$  are the hydroelectric output and fossil-electric output in period  $i$ , respectively,  $C_i(q_i^T)$  is the cost of the fossil-fuelled unit to produce  $q_i^T$ , and  $\bar{q}^H$  is the total hydro energy available for the two periods. Assume period 1 is the off-peak with lower demand, and period 2 the peak demand period. Thus,  $A_2$  is greater than  $A_1$ , and  $Q_2^*$  is greater than  $Q_1^*$ .

Similar to the one period model, there can be three possible outcomes: the equilibrium price can be lower than, higher than, or equal to the price cap, depending on market conditions, cost and regulatory parameters. But the monopolist in the two-period model has more flexibility, as it now has two separate markets rather than one (since we do not model the intertemporal substitution of consumption, the peak and off-peak periods are essentially two separate markets).

*Case 1*

Assume that the water supply is large enough to make it optimal for the monopolist to use it in both periods<sup>23</sup>. Without the rebate mechanism, or when it is to the monopolist's benefit to bid below the price cap (for example, the price cap is greater than the unconstrained monopoly price), the rebate will be equal to zero. To distinguish this from other scenarios, denote this case as Case 1. Then the monopolist's problem becomes the standard unconstrained monopoly problem. In this case, the first order conditions (after substituting the water constraint into the objective function to eliminate the variable  $q_2^H$ ) for the monopoly case are:

$$\frac{\partial \pi_m}{\partial q_1^H} = A_1 - A_2 + 2B\bar{q}^H - 4Bq_1^H - 2Bq_1^T + 2Bq_2^T = 0 \quad (3.3)$$

$$\frac{\partial \pi_m}{\partial q_1^T} = A_1 - b - 2Bq_1^H - (2B + c)q_1^T = 0 \quad (3.4)$$

$$\frac{\partial \pi_m}{\partial q_2^T} = A_2 - b - 2B\bar{q}^H + 2Bq_1^H - (2B + c)q_2^T = 0 \quad (3.5)$$

It is straightforward to derive the simple monopoly solutions by solving for the three unknown variables. The equilibrium is as follows:

$$q_{m1}^H = \frac{1}{4B} (A_1 - A_2 + 2B\bar{q}^H) \quad (3.6)$$

$$q_{m2}^H = \frac{1}{4B} (A_2 - A_1 + 2B\bar{q}^H) \quad (3.7)$$

$$q_{m1}^T = \frac{1}{2(2B+c)} (A_1 + A_2 - 2b - 2B\bar{q}^H) \quad (3.8)$$

$$q_{m2}^T = \frac{1}{2(2B+c)} (A_1 + A_2 - 2b - 2B\bar{q}^H) \quad (3.9)$$

$$P_{m1} = \frac{1}{4(2B+c)} (4A_1B + 3A_1c + 4Bb + A_2c - 2Bc\bar{q}^H) \quad (3.10)$$

$$P_{m2} = \frac{1}{4(2B+c)} (4A_2B + 3A_2c + 4Bb + A_1c - 2Bc\bar{q}^H) \quad (3.11)$$

where the subscript  $m$  denotes an unconstrained monopoly.

---

<sup>23</sup> If there is little water, the optimal solution is obvious: allocate all water in the period where the price is highest.

It can be seen that equation (3.8) is exactly same as equation (3.9), implying that the monopolist generates electricity until the marginal costs of thermal unit in both periods are equal. Thus a rational monopolist will produce an equal amount of thermal electricity in each period<sup>24</sup>. In other words, the monopolist will generate electricity until marginal revenue from thermal output is equal in the two periods.

Because  $A_2 > A_1$ , hydro output will be higher in the peak period than in the off-peak period. The shadow price of water (the Lagrange multiplier of the water constraint) is:

$$\lambda_m = \frac{1}{2(2B+c)} (A_1c + A_2c + 4Bb - 2Bc\bar{q}^H) \quad (3.12)$$

The shadow price of hydro is an increasing function of the demand level, and a decreasing function of the available water reserve, the efficiency level of the fossil unit and the elasticity of demand. Recall that we are assuming no variable cost for a hydro generating unit, thus the shadow price is the opportunity cost of the water resource. In equilibrium, the opportunity cost of water in both periods should be equal, and equal to the marginal cost of the fossil unit.

### *Case 2*

In the case where the price cap is so low that it is to the monopolist's benefit to bid above the price cap (denoted as Case 2), the first order conditions for the problem are:

---

<sup>24</sup> This resembles the case in which a multi-plant monopolist equalizes marginal cost among all plants.

$$\frac{\partial \pi_m}{\partial q_1^H} = A_1 - A_2 + 2B\bar{q}^H + BQ_1^* - BQ_2^* - 4Bq_1^H - 2Bq_1^T + 2Bq_2^T = 0 \quad (3.13)$$

$$\frac{\partial \pi_m}{\partial q_1^T} = A_1 - b - 2Bq_1^H + BQ_1^* - (2B + c)q_1^T = 0 \quad (3.14)$$

$$\frac{\partial \pi_m}{\partial q_2^T} = A_2 - b - 2B\bar{q}^H + BQ_2^* + 2Bq_1^H - (2B + c)q_2^T = 0 \quad (3.15)$$

Compared to the FOCs of a simple monopoly (Case 1), each of the first order conditions has an additional term associated with the rebate mechanism. From equations (3.14) and (3.15), one can see that the marginal revenue net of rebate is now higher than when the rebate mechanism is not in place, implying that the *MR* curve is shifted up (as shown in the one period model), and thus thermal output will be higher. Solving the FOCs yields the following equilibrium output and prices:

$$q_{b1}^H = \frac{1}{4B} [A_1 - A_2 + 2B\bar{q}^H + BQ_1^* - BQ_2^*] \quad (3.16)$$

$$q_{b2}^H = \frac{1}{4B} [A_2 - A_1 + 2B\bar{q}^H - BQ_1^* + BQ_2^*] \quad (3.17)$$

$$q_{b1}^T = \frac{1}{2(2B+c)} [A_1 + A_2 - 2b - 2B\bar{q}^H + BQ_1^* + BQ_2^*] \quad (3.18)$$

$$q_{b2}^T = \frac{1}{2(2B+c)} [A_1 + A_2 - 2b - 2B\bar{q}^H + BQ_1^* + BQ_2^*] \quad (3.19)$$

$$P_{b1} = \frac{1}{4(2B+c)} [4A_1B + 3A_1c + 4Bb + A_2c - 2Bc\bar{q}^H - B(4B+c)Q_1^* + BcQ_2^*] \quad (3.20)$$

$$P_{b2} = \frac{1}{4(2B+c)} [4A_2B + 3A_2c + 4Bb + A_1c - 2Bc\bar{q}^H + BcQ_1^* - B(4B+c)Q_2^*] \quad (3.21)$$

Similar to Case 1, the constrained monopolist produces the same amount of thermal electricity in both the peak and off-peak periods. Note that because  $A_2 > A_1$  and  $Q_2^* > Q_1^*$ , the firm is now using more hydro to shave the peak demand, compared to Case 1 where the rebate mechanism is not binding. The shadow price of water now is:

$$\lambda_b = \frac{1}{2(2B+c)} (A_1c + A_2c + 4Bb - 2Bc\bar{q}^H + BcQ_1^* + BcQ_2^*) \quad (3.22)$$

which is again equal to the marginal cost of the thermal unit, but greater than the shadow price in Case 1, implying that the water becomes more valuable in Case 2 as more thermal output is required.

Again, the shadow price depends positively on the demand level and the  $CRQ$ 's for both periods, and negatively on the available water stock and the elasticity of demand. As the  $CRQ$ 's increase, the water resource becomes more valuable to the monopolist. The intuition here is that the MPMA induces the monopolist to produce more than it would otherwise, and thus more high cost thermal electricity. To minimize the cost of the requisite additional thermal generation, the monopolist allocates more of its water resource to the high demand period.

Note that in Case 2 the peak price is not necessarily greater than the off-peak price – a different result from an unconstrained monopoly. To see this, subtract (3.20) from (3.21) to obtain  $P_{b2} - P_{b1} = \frac{1}{2}[A_2 - A_1 + B(Q_1^* - Q_2^*)]$ , which can be positive or negative, depending mainly on how large the  $CRQ$ 's are. For example, if  $Q_2^*$  is far larger than  $Q_1^*$  due to significant over-forecast of peak demand, the peak price  $P_{b2}$  could be smaller than the off-peak price  $P_{b1}$  (similar to “price reversal” as documented by Bailey and White, 1974).

The binding price cap implies  $P_{b1}^* Q_1^* + P_{b2}^* Q_2^* > P_{cap} (Q_1^* + Q_2^*)$ . Substituting both equation (3.20) and (3.21) into the inequality yields:

$$P_{cap} < AP_m - \frac{B((Q_1^*)^2 + (Q_2^*)^2)}{2(Q_1^* + Q_2^*)} + \frac{Bc(Q_1^* + Q_2^*)}{4(2B+c)} \quad (3.23)$$

where  $AP_m$  denotes the  $CRQ$ -weighted unconstrained monopoly price. The last two terms on the right-hand side can be combined as  $-\frac{B[4B(Q_1^{*2} + Q_2^{*2}) + c(Q_1^* - Q_2^*)^2]}{4(2B+c)(Q_1^* + Q_2^*)}$ , which must be

negative. That is, a binding price cap must be lower than the unconstrained monopoly price by some amount, which depends on the  $CRQ$ 's, the slope of the demand curve, and the slope of the marginal cost curve.

It is straightforward to show that this inequality is more likely satisfied with a low price cap, less efficient thermal generation (a large  $b$  and/or  $c$ ), high demand levels (large  $A_1$  and/or  $A_2$ ), a less elastic demand (low  $B$ ), and/or a small stock of available water (low  $\bar{q}^H$ ). For example, a high demand level and/or a low elasticity of demand, less efficient generation and/or a small stock of the water resource could significantly increase the market clearing price so that the average price could be greater than the price cap.

---

Example 3.1: Assume the demand functions for off-peak and peak are:  $P_1 = 80 - q_1$  and  $P_2 = 120 - q_2$ , respectively. The monopolist's cost function of its thermal unit is  $C_i(q_i^T) = 3 + 2q_i^T + 0.1(q_i^T)^2$ . The maximum available water  $\bar{q}^H$  is 60MW and the CRQ is 60MW for period 1 and 90MW for period 2. The equilibria for the two scenarios are listed in the following table. Note for simplicity of calculation, the coefficients of the cost function here are different from typical values in the industry as mentioned before.

One can see in the constrained case, there is more hydro production on-peak, and more thermal production both off-peak and on-peak. If the price cap is chosen greater than or equal to \$54.73/MW-h, the best strategy for the monopolist is to set the weighted average price to be \$54.73/MW-h. In contrast, if the price cap is lower than \$19.14/MW-h, the monopolist will charge a weighted average price \$19.14/MW-h, which will be higher than the price cap. In the former case, the monopolist earns a profit of \$5050.36; in the latter, it makes  $150P_{\text{cap}} - 489.82$ , which is positively related to the price cap, with a maximum \$2381.18 at a price cap of \$19.14.

It's apparent that the monopolist will bid the price cap if the price cap is between \$19.14 and \$54.73. Note that in these cases the CRQ's are chosen before the price cap is chosen. However, if the price cap is chosen first and is lower than \$54.73, the policy choice will be to set respective CRQ's so that the firm decides to produce at or above the price cap, which becomes our Example 2.

**Table 3.1: Comparison of an Unconstrained Monopolist and a Monopolist Who Bids Above the Price Cap**

	Unconstrained Monopoly (Case 1)	Constrained Monopoly (Case 2)
$q_1^H$	20MW	12.5MW
$q_2^H$	40MW	47.5MW
$q_1^T$ and $q_2^T$	17.27MW	51.36MW
$P_1$	\$42.73	\$16.14
$P_2$	\$62.73	\$21.14
CRQ weighted P	\$54.73	\$19.14
condition of $P_{cap}$	$\geq \$54.73$	$\leq \$19.14$
Profit	\$5050.36	$150P_{cap} - 489.82$

### Case 3

In the third case the monopolist bids the price cap, i.e.  $(P_1\beta + P_2)/(1 + \beta) = P_{cap}$ , where  $\beta$  denotes the ratio of CRQ for off-peak to CRQ for peak, there will be no refund to consumers. As long as  $P_1 \neq P_2$ ,  $\beta$  plays an important role in the monopolist's production decision. Substituting the demand function into the constraints yields  $(q_2^H + q_2^T) = \alpha - \beta(q_1^H + q_1^T)$ , where  $\alpha = \frac{1}{B}[A_1\beta + A_2 - P_{cap}(1 + \beta)]$ .

Therefore the firm's problem can be written as:

$$MAX : \pi_r = \sum(P_i p_i - C_i(q_i)) \quad (3.24)$$

$$s.t.: \quad (1) P_i = A_i - B(q_i^H + q_i^T) \quad \forall i = 1, 2$$

$$(2) C_i(q_i^T) = a + bq_i^T + \frac{1}{2}c(q_i^T)^2$$

$$(3) (q_2^H + q_2^T) = \alpha - \beta(q_1^H + q_1^T)$$

$$(4) \alpha = \frac{1}{B}[A_1\beta + A_2 - P_{cap}(1 + \beta)]$$

$$(5) \beta = Q_1^* / Q_2^*$$

$$(6) \sum q_i^H = \bar{q}^H$$

Substituting the constraints (1)-(6) into the objective function and taking the partial derivatives with respect to  $q_1^H$  and  $q_1^T$  respectively yields the FOC's:

$$\frac{\partial \pi_r}{\partial q_1^H} = A_1 - A_2\beta + 2B\alpha\beta - b(1 - \beta) - c(1 - \beta)(\alpha - \bar{q}^H) - [2B + 2B\beta^2 + c(1 - \beta)^2]q_1^H - [2B + 2B\beta^2 - c\beta(1 - \beta)]q_1^T = 0 \quad (3.25)$$

$$\frac{\partial \pi_r}{\partial q_1^T} = A_1 - A_2\beta + 2B\alpha\beta - b(1 - \beta) + c\beta(\alpha - \bar{q}^H) - [2B + 2B\beta^2 - c\beta(1 - \beta)]q_1^H - [2B + 2B\beta^2 + c(1 + \beta^2)]q_1^T = 0 \quad (3.26)$$

These FOC's are very complicated and a comparison of them to the FOC's of other cases is impossible. However, one obvious observation in this case is that the ratio of  $CRQ$ 's matters, rather than the levels of the  $CRQ$ 's. Nevertheless, in later sections we will show the  $CRQ$ 's must be greater than certain values.

Solving this problem results in the following equilibrium:

$$q_{r1}^H = \frac{(1+\beta)(A_1 - A_2\beta + 2B\alpha\beta - b + b\beta) + (\alpha - \bar{q}^H)(\beta c - 2B\beta^2 - 2B - c)}{w} \quad (3.27)$$

$$q_{r1}^T = \frac{(1-\beta)(A_1 - A_2\beta + 2B\alpha\beta - b + b\beta) + 2B(\alpha - \bar{q}^H)(1 + \beta^2)}{w} \quad (3.28)$$

$$q_{r2}^H = \frac{-(1+\beta)(A_1 - A_2\beta - b + b\beta) + \alpha(-2B\beta - c\beta + 2B + c) + \bar{q}^H(c\beta^2 - \beta c + 2B + 2B\beta^2)}{w} \quad (3.29)$$

$$q_{r2}^T = \frac{(1-\beta)(A_1 - A_2\beta + 2B\alpha\beta - b + b\beta) + 2B(\alpha - \bar{q}^H)(1 + \beta^2)}{w} \quad (3.30)$$

$$P_{r1} = \frac{1}{w} \{ A_1[w - 2B] - 2B[-A_2\beta + 2B\alpha\beta - b + b\beta] + B\alpha(1 - \beta)(\alpha - \bar{q}^H) \} \quad (3.31)$$

$$P_{r2} = \frac{1}{w} \{ 2B\beta[A_1 + 2B\alpha\beta - b + b\beta] + A_2[w - 2B\beta^2] - B\alpha(4B + 4B\beta^2 + c - c\beta) + Bc\bar{q}^H(1 - \beta) \} \quad (3.32)$$

where  $w = 4B(1 + \beta^2) + c(1 - \beta)^2$ .

From the equilibrium equations alone, it is also impossible to compare these quantities to those of Case 1 or 2. It is shown in Example 3.2 that the thermal outputs in Case 3 are the highest amongst all three cases, given a price cap that is below the unconstrained monopoly price, which in turn implies the shadow price of water is the highest. Thus the peak hydro output must be the highest in Case 3.

One implication of the model is that there is no need for peaking fossil capacity in the thermal plant because of the flexibility of the hydroelectric facility. Rather, the hydro unit takes the full responsibility for shaving the peak demand.

The difference between Case 2 and Case 3 can be illustrated with the following numerical example.

---

Example 3.2: Assume the demand functions for off-peak and peak are  $P_1 = 80 - q_1$  and  $P_2 = 120 - q_2$ , respectively. The monopolist's cost function of

thermal unit is  $C_i(q_i^T) = 3 + 2q_i^T + 0.1(q_i^T)^2$ . The maximum available water  $\bar{q}^H$  is 60MW and the price cap is \$15/MWh. The CRQ's are to be determined. The equilibria are listed in the following table.

**Table 3.2: Comparison of a Monopolist Who Bids Above the Price Cap and a Monopolist Who Bids Below the Price Cap**

	Case 2	Case 3
$q_1^H$	$\frac{1}{4}(80 + Q_1^* - Q_2^*)$	$\frac{145\beta^2 + 36\beta - 21}{4.2 - 0.4\beta + 4.2\beta^2}$
$q_2^H$	$\frac{1}{4}(160 - Q_1^* + Q_2^*)$	$\frac{107\beta^2 - 60\beta + 273}{4.2 - 0.4\beta + 4.2\beta^2}$
$q_1^T$ and $q_2^T$	$\frac{1}{4.4}(76 + Q_1^* + Q_2^*)$	$\frac{128\beta^2 + 144\beta + 168}{4.2 - 0.4\beta + 4.2\beta^2}$
$P_1$	$\frac{1}{8.8}(376 - 4.2Q_1^* + 0.2Q_2^*)$	$\frac{189 - 212\beta + 63\beta^2}{4.2 - 0.4\beta + 4.2\beta^2}$
$P_2$	$\frac{1}{8.8}(552 - 4.2Q_2^* + 0.2Q_1^*)$	$\frac{63 - 128\beta + 269\beta^2}{4.2 - 0.4\beta + 4.2\beta^2}$
CRQ weighted price	$\frac{376Q_1^* + 552Q_2^* + 0.4Q_1^*Q_2^* - 4.2((Q_1^*)^2 + (Q_2^*)^2)}{8.8(Q_1^* + Q_2^*)}$	15
conditions for $Q_1^*$ and $Q_2^*$	<ol style="list-style-type: none"> <li>1. <math>\frac{376Q_1^* + 552Q_2^* + 0.4Q_1^*Q_2^* - 4.2((Q_1^*)^2 + (Q_2^*)^2)}{8.8(Q_1^* + Q_2^*)} &gt; 15</math></li> <li>2. <math>q_1^H, q_2^H, q_1^T, q_2^T, P_1, P_2 \geq 0</math></li> </ol>	<ol style="list-style-type: none"> <li>1. <math>\frac{376Q_1^* + 552Q_2^* + 0.4Q_1^*Q_2^* - 4.2((Q_1^*)^2 + (Q_2^*)^2)}{8.8(Q_1^* + Q_2^*)} &lt; 15</math></li> <li>2. <math>q_1^H, q_2^H, q_1^T, q_2^T, P_1, P_2 \geq 0</math></li> </ol>

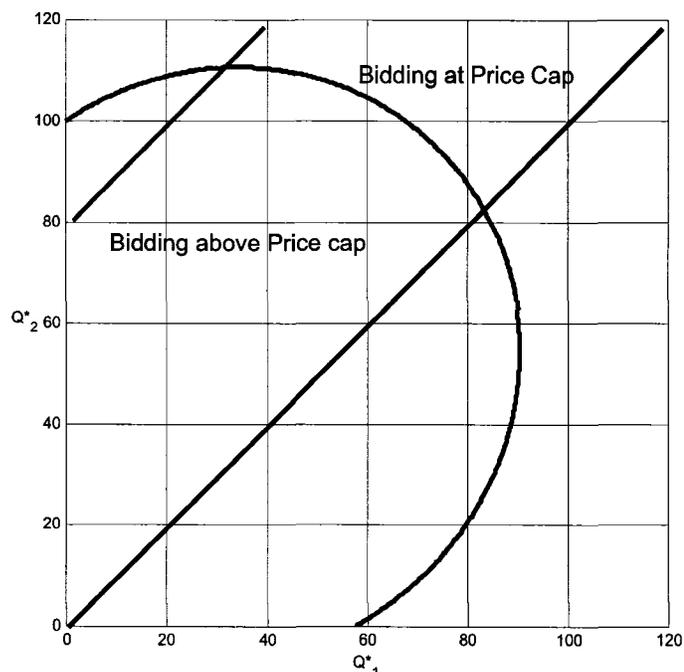
The first condition for the CRQ's (in the last row in the table) in Case 2 is to ensure that the policy induces the monopolist to bid price higher than the cap, the second condition is to preclude scenarios with negative equilibrium prices or quantities. The first condition for bidding at the price cap (Case 3) is derived simply because if the firm is not bidding above the cap it must be bidding at the cap given that the price cap (\$15) is much lower than the unconstrained monopoly price (\$54.73). The second condition for Case 3 is the same as the case of bidding above the cap (Case 2).

The following example, Figure 3.1 illustrates the conditions for the monopolist to bid at or above the price cap under the above settings. The curve consists of all solutions that could make the weighted price equal to the price cap based on the equality Condition 1 in the above table. Because  $Q_2^*$  is greater than  $Q_1^*$  by assumption, the solution to each case must be in the upper left half of the diagram i.e. the large triangular area. Also, if the prices and quantities are non-negative, the set of solutions then must be below the cut line in the upper left corner (which is derived simply by making  $q_1^H$  greater than or equal to zero<sup>25</sup>). Therefore the area between the two slash lines and below the arc curve is the set of CRQ's that induce the monopolist to bid above the price cap. The area outside the arc curve and between the two slash lines is the set of CRQ's that induce the monopolist to bid at the price cap.

<sup>25</sup> It is possible that there is not enough water to generate electricity in both periods if there is a minimum level of production. In that case, all water will be used in period 2 when the opportunity cost is higher, However, this is outside our scope of interest.

Of course, it is possible that the CRQ in one period is chosen too high, for example, the set of CRQ's located above the upper left slash line (the small triangular area), such that all hydro will be allocated to this single period. However, we won't discuss these special cases in detail because our main interest is the allocation of water.

*Example Figure 3.1 Conditions of CRQs Given a Price Cap*

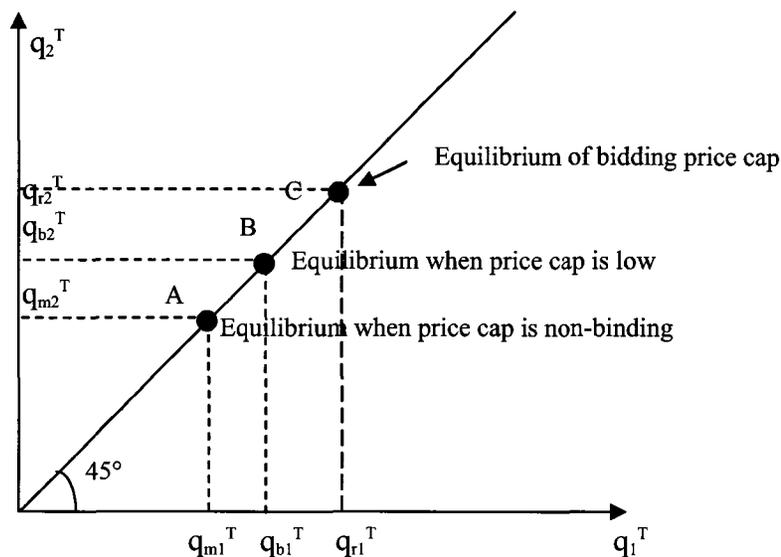


One can see that to induce the firm to bid at the pre-determined price cap, the policy maker has to set relatively higher CRQ's. In this case the monopolist will generate in general more electricity in aggregate, if not in each period, compared to the case of bidding above the price cap. Given that the hydro reserve is constant, this means that total thermal electricity production is higher, which in turn implies the hydro production is higher in the peak period in Case 3 than in Case 2.

From the above numerical example, we can infer that the thermal output is in general highest peak period hydro output is highest in Case 3.

Figure 3.1 below illustrates the possible production points of thermal generation: all output points are on the  $45^\circ$  line. From the above examples, we know that thermal electricity production in Case 2 (point  $B$ ) must be higher than that in Case 1 (point  $A$ ). As demonstrated in Example 3.2, the solutions to Case 2 are multiple, given any price cap. Therefore point  $B$  can be any point between  $A$  and  $C$  along the  $45^\circ$  ray. Similarly, the solutions to Case 3 are multiple as well, and point  $C$  can be any point on the  $45^\circ$  ray beyond point  $B$ . From the perspective of the social optimum, Case 3 should be the candidate because the outputs are the highest among all three cases.

**Figure 3.1: Thermal Production in Equilibrium under the MPMA**



An increase in the water resource (e.g. more rainfall) has a different impact on hydro and thermal production in the three cases. In Cases 1 and 2, the monopolist will evenly split its additional water between two periods (because the coefficient of  $\bar{q}^H$  in equations (3.6) and (3.16) is the same as in equations (3.7) and (3.17)). Total energy production in each period will increase although thermal production will drop. This is equivalent to a

rightward shift of the marginal cost curve in both periods, and thus the equilibrium will move to the right. Consequently, the prices in the peak and off-peak periods will drop by the same amount (an apparent result of linear demand and linear marginal cost).

In Case 3, in contrast, an increase in the water resource has an asymmetric effect on both the peak and off-peak hydro production and prices. Equation (3.27) shows that an increase in the available water resource will lead to an increase in hydro output in period 1 by  $\frac{2B+2B\beta^2+c-c\beta}{4B+4B\beta^2+c(1-\beta)^2}$ , which is greater than the increase in period 2,  $\frac{2B+2B\beta^2+c\beta^2-c\beta}{4B+4B\beta^2+c(1-\beta)^2}$  (because  $\beta < 1$  by assumption). This production allocation implies that the marginal revenue of water off-peak is lower than on-peak if the monopolist is bidding at the price cap.

Similar to Cases 1 and 2, the monopolist's total output here will increase in each period if the off-peak *CRQ* is lower than the peak *CRQ*. A special case occurs when the *CRQ*'s are equal in both periods, the monopolist will lower its thermal output by the same amount as the increases in hydro output, and as a result, there will be no impact on the equilibrium prices, following equations (3.29) and (3.30). However if the increase in the water supply is extremely high, thus shifting the MC curve much further to the right, the firm might be better off producing at a price below the cap.

On the other hand, if the thermal unit has a constant marginal cost, it is interesting to see that the market equilibrium prices are not affected by the change in the availability of water, which means that the increase in hydro production is completely offset by the decrease in thermal output.

A change in input cost ( $b$  and/or  $c$ ), for example, because of a change in fuel price and/or the aging of a generating unit, has different impacts on the equilibria of the three cases. In Cases 1 and 2, the cost parameters do not enter the function of hydro production, and thus have no impact on the water allocation. But a change in the cost function does affect thermal production. For example, an increase in the fuel price, which will increase  $b$  (and  $c$ ), will lead to lower thermal output, while a fall in the fuel price will result in an increase in thermal production. The change in cost, however, will affect the hydro production in Case 3. According to equation (3.27), an increase in  $b$  or  $c$  will lead to lower hydro production in off-peak, and therefore higher hydro output on-peak. The intuition here is that the cost increase will translate into an increase in the monopolist's bid price, which would put it above the price cap and reduce its profits. To adhere to the price cap, the monopolist has to reduce its peak bid price, which has a greater weight in the average price, and thus increase hydro production on-peak

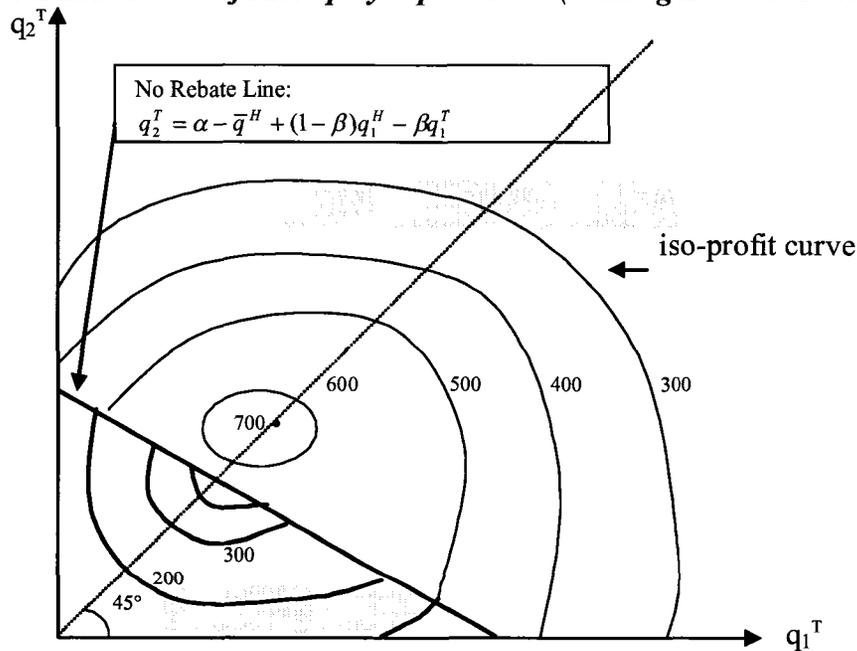
### **3.2 Non-uniqueness of Policies to Induce the Monopolist to Bid at a Specified Price Cap**

*Example 3.2* illustrates a numerical example of the non-uniqueness of the optimal policy. Here we will demonstrate it in a more general way.

In the situation of unconstrained monopoly, the average equilibrium price is lower than the stipulated price cap. Figure 3.2 illustrates a hypothetical case where the price cap is so high that even the monopoly price is less than the cap. There are two areas in the

diagram: the non-binding area, where the bidding price is lower than the price cap, and the binding area, where the bidding price is higher than the price cap. The two areas are separated by a “no rebate line”, which is derived by substituting the demand function into the rebate formula (the constraint (3) in problem (3.24) which establishes that the weighted price equals the bid cap, hence no rebate). This line can shift up and down, depending on the availability of water, the  $CRQ$ 's, the price cap, and the demand level. Its slope equals  $\beta$ , the ratio of the  $CRQ$ 's. An increase in the price cap will shift the rebate line down, and an increase in any  $CRQ$  will shift the line up. A relative increase in first period's  $CRQ$  will rotate the rebate line clockwise, while a relative increase in the second period's  $CRQ$  will turn it counterclockwise.

**Figure 3.2 Illustration of Monopoly Equilibrium (Bidding Below the Price Cap)**



The profit function in the non-binding area can be expressed as:

$$\pi_{non-binding} = -2a + (A_1 - b)(q_1^H + q_1^T) + (A_2 - b)(\bar{q}^H - q_1^H + q_2^T) - (B + \frac{1}{2}c)[(q_1^T)^2 + (q_2^T)^2] \quad (3.33)$$

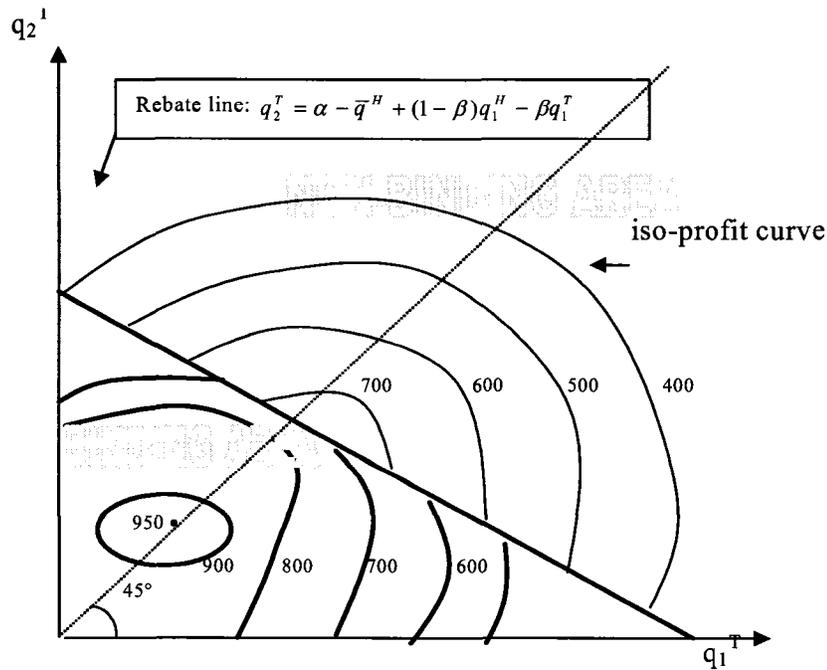
and in the binding area

$$\pi_{binding} = P_{cap}(Q_1^* + Q_2^*) - 2a - A_1Q_1^* - A_2Q_2^* + (A_1 - b + BQ_1^*)(q_1^H + q_1^T) + (A_2 - b + BQ_2^*)(\bar{q}^H - q_1^H + q_2^T) - (B + \frac{1}{2}c)[(q_1^T)^2 + (q_2^T)^2] \quad (3.34)$$

Given an output level of  $q_1^H$  ( $\leq \bar{q}^H$ ), the profit functions are quadratic in both  $q_1^T$  and  $q_2^T$ , and thus the iso-profit curves have the general shapes shown in the following three graphs. Their positions and exact shapes vary, depending on demand factors (i.e. levels and slopes), cost parameters, water allocation between periods, the price cap, and the  $CRQ$ 's (when the price cap is binding). The iso-profit curve must be continuous in each case according to the equations (3.33) and (3.34).

Generally speaking, if the monopolist produces a small amount of electricity in both periods, or a very small amount of electricity in either period, the average price would be higher than the price cap, and thus the firm has to pay back to consumers according to the rebate formula. The combination of thermal outputs resulting in a rebate is represented by the area below the “no-rebate line”. In contrast, if the firm produces a large amount of electricity, the average price will be lower than the price cap, and as a result the firm doesn't have to pay back to consumers. This situation falls in the area above the “no-rebate line”.

**Figure 3.3 Illustration of Monopoly Equilibrium (Bidding Above the Price Cap)**



In Case 1 (Figure 3.2), there is a unique local maximum in the non-binding area according to equation (3.8) and (3.9), which is also globally optimal because the profit in the binding area can never be greater than the profit in non-binding area.

In the meantime, the shape and position of the iso-profit curve in the binding area depends on both the price cap and the  $CRQs$ . An increase in the price cap will increase the monopolist's profit (because of a lower rebate) but doesn't change the shape of the iso-profit curves, while an increase in  $CRQs$  will lower its profit and change the shape of the iso-profit curve. It is apparent that profit is lower under the price cap (MPMA) given all parameters unchanged.

In short, an increase (decrease) in the price cap can shift the no rebate line down (up), and move all iso-profit curves up (down) in the binding area. An increase (decrease) in the relative  $CRQ$  can make the rebate line steeper (flatter), and shift it out (inward), and change the shape of iso-profit curves and lower (increase) profit in the binding area.

**Figure 3.4 Illustration of Monopoly Equilibrium (Bidding at the Price Cap)**

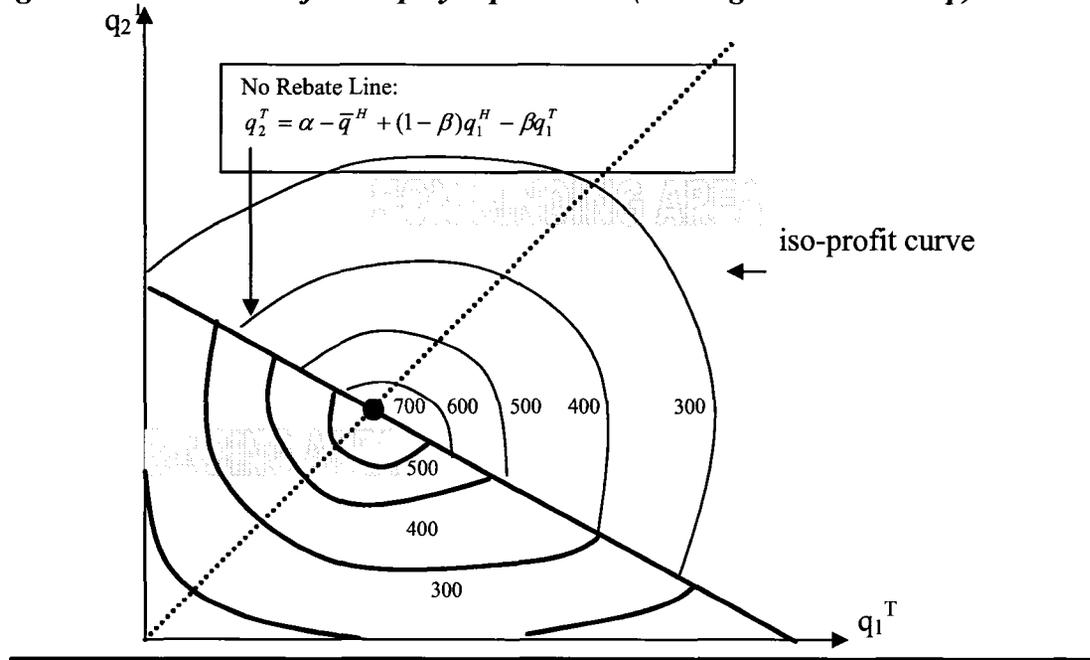


Figure 3.3 illustrates Case 2. The monopolist can attain the highest profit of \$950 in the “binding area”; thus the monopolist will prefer producing at low output level (bidding above the cap) and paying a rebate. This is the case when the price cap and  $CRQ$ s are low enough (i.e. the “no-rebate line” is far enough away from the origin) that it is always profitable for the firm to produce in the “binding area”. This graph also explains why the firm might bid above the price cap because along the “no rebate line” the profit is always lower than that at the constrained production point. It is intuitive that to induce the firm

to bid at the price cap, the price cap and  $CRQs$  must be carefully chosen such that the maximum profit in the binding area is lower than that on the “no rebate line”.

Case 3 arises when the price cap is sufficiently low and  $CRQs$  are sufficiently large such that the maximum profit point is on the “no-rebate line” (Figure 3.4). If the purpose of the MPMA is to induce the monopolist to bid at the price cap, this would be the desired outcome, and thus the task of the market designers is to find the “correct” combination of price cap and  $CRQs$  to ensure that the monopolist bids at the cap.

To ensure that the monopolist bids at the price cap, there are two necessary conditions. First, the price cap must be lower than the unconstrained monopoly price, i.e. the equilibrium cannot be Case 1. Second, the profit of bidding at the price cap must be higher than the profit of bidding above the price cap. These two conditions can be expressed as:

$$\pi_r > \pi_b \quad (3.35)$$

$$P_{m1}\beta + P_{m2} > (1 + \beta)P_{cap} \quad (3.36)$$

Substituting the equilibrium conditions of Case 3 into the above necessary conditions yields the following inequality:

$$AP_m - \frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)} - \varepsilon < P_{cap} < AP_m - \frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)} + \varepsilon$$

*or*  $< AP_m$  , *whichever is less* (3.37)

where  $AP_m$  is the  $CRQ$ -weighted unconstrained monopoly price,  $\frac{P_{m1}Q_1^* + P_{m2}Q_2^*}{Q_1^* + Q_2^*}$ ,

$w = 4B(1 + \beta^2) + c(1 - \beta)^2$ , and  $\varepsilon$  is a positive term and greater than  $\frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)}$ , which is

explained later. It can be shown that equation (3.37) is equivalent to

$$AP_m - \frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)} - \varepsilon < P_{cap} < AP_m \quad (3.37')$$

To see this, consider the case where the price cap is just slightly lower than the monopoly price and the  $CRQ$  in each period is any number that is greater than the monopoly output.

Because the price cap is sufficiently low, bidding above the price cap won't be as

profitable as bidding at the price cap<sup>26</sup>. On the other hand, bidding below the price cap is

not profitable either. The intuition here is that when the price is below the unconstrained monopoly price, profit decreases as output increases. Therefore the only profit

maximizing strategy for the monopolist is to bid at the price cap. This implies that the

profit difference  $\pi_r - \pi_b$  is positive within the small range of prices below the monopoly

price. Therefore  $AP_m - \frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)} + \varepsilon$  must be greater than  $AP_m$ . Figure 3.5

depicts this relationship between the price cap, the  $CRQ$  and the monopoly price.

$\varepsilon$  is a complicated function of a number of factors, and can be expressed as:

---

<sup>26</sup> Recall in a single period model a price cap that is slightly lower than the monopoly price indicates that the disjointed point of the marginal revenue curve is just slightly to the right of the monopoly quantity. As a result the marginal cost will not intersect the marginal revenue curve at its upper part, and thus the bidding above the price cap would not increase profits.

$$\varepsilon = \left\{ \frac{w}{2B(2B+c)} \left[ \frac{2B[(2B+c)(\beta q_{m1} + q_{m2}) - 2B\bar{q}^H]^2}{(2B+c)w} - \frac{A_1 q_{m1} + A_2 q_{m2}}{2} + \frac{b(A_1 + A_2 - 2b)}{2(2B+c)} \right] + \frac{(A_1 - A_2\beta - b + b\beta + 2B\beta\bar{q}^H)^2}{w} - \frac{B(\bar{q}^H)^2(4B+c)}{2(2B+c)} + \frac{\bar{q}^H(8A_2B + 3A_2c - A_1c - 8Bb + 4BcQ_2^*)}{4(2B+c)} \right\}^{\frac{1}{2}}$$

which is positive and greater than  $\frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)}$ , given  $AP_m - \frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)} + \varepsilon$  greater than  $AP_m$ .

A rule of thumb to find an appropriate policy combination of price cap and  $CRQ$ 's is to choose an equation that satisfies condition (3.37'), such as

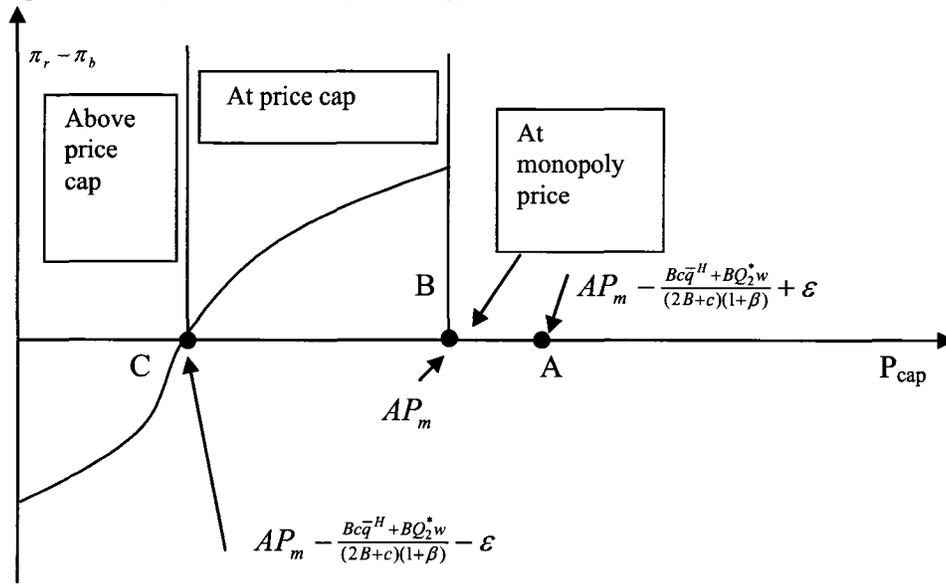
$$P_{cap} = AP_m - \frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)} \quad (3.38)$$

Any policy combination that satisfies the above equation will induce the monopolist to bid at the price cap. The necessary condition has the following characteristics: the higher the demand level or the less elastic the demand, the higher the price cap should be; the greater is the supply of water, the lower price cap should be; and the greater is the marginal cost of thermal generation, the higher the price cap should be. The equation also depicts the relationship between the two policy variables: the lower the price cap, the higher the  $CRQ$ 's, and *vice versa*.

The strategies of the monopoly under different price caps are illustrated in Figure 3.5. If the price cap is lower than  $AP_m - \frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)} - \varepsilon$ , it is always profitable for the monopoly to bid above the cap. If the price cap is greater than the unconstrained monopoly price,

the monopolist will always bid at the monopoly price. If the price cap is in between, the monopolist will bid at the cap. Thus the policy implication is that there are many price cap and  $CRQ$  combinations that can induce the monopolist to bid at the desired price level, whether or not the price cap is equal to the social optimum price. Therefore we have Proposition 1.

**Figure 3.5 Optimal Bidding Strategy**



**Proposition 1 (effectiveness of the price cap)** There are multiple policy choices to induce the monopolist to bid at the pre-specified price cap. For any price cap  $P_{cap}$ , a

combination of  $Q_1^*$  and  $Q_2^*$  that satisfies  $AP_m - \frac{4Bc\bar{q}^H + BQ_2^*w}{4(2B+c)(1+\beta)} < P_{cap}$  (where

$w = 4B(1 + \beta^2) + c(1 - \beta)^2$ ) can induce the regulated monopoly firm to bid at the price cap.

### 3.3 Welfare Implications of the MPMA

As was shown in the case of the one period model, the MPMA is welfare improving (also see Chute, 2001), which is also implied by Proposition 1. The MPMA is more beneficial, however, to peak consumers. This can be shown by comparing equations (3.10) and (3.11) to (3.20) and (3.21), respectively. Because  $[-(4B^2 + Bc)Q_1^* + BcQ_2^*]$  (in equation (3.20)) is less than  $[BcQ_1^* - (4B^2 + Bc)Q_2^*]$  (in equation (3.21)) as a result of  $Q_1^* < Q_2^*$ , the peak price will drop more than the off-peak price.

In Case 3 where the price cap is the equilibrium price, thermal output is the highest (Figure 3.1), and hydro production will shift most from off-peak to peak (see Proposition 3 below). Therefore, the total peak output will increase relatively more than the off-peak output, the peak price will fall more, and consequently the peak consumers benefit more than the off-peak consumers.

Thus we have Proposition 2.

**Proposition 2 (welfare effect)** *The MPMA price cap is welfare improving as long as the price cap is lower than the CRQ-weighted unconstrained monopoly price  $AP_m$ . Peak consumers benefit more from the imposition of the MPMA.*

### 3.4 Water Allocation

First, let's compare the water allocation in Cases 1 and 2. By comparing equation (3.6) and (3.7) to (3.16) and (3.17), respectively, one finds that  $q_{b1}^H < q_{m1}^H$ , while  $q_{b2}^H > q_{m2}^H$ , because  $Q_1^* < Q_2^*$ . Therefore the monopolist shifts its hydraulic power production from off-peak to peak when it bids above the price cap, compared to an unconstrained monopolist.

It is not straightforward to compare hydro output in Cases 3 and 1. It is, however, possible to infer such a conclusion as was shown in *Example 3.2*. Recall that the unconstrained monopoly is the worst case in terms of social welfare among all three scenarios. Given that Case 3 mitigates the monopolist's market power, peak hydro production thus must be increased compared to Case 1, and accordingly the off-peak hydro production is lower. Bushnell (2000) shows that hydrothermal producers in a Cournot oligopoly with a competitive fringe have an incentive to shift hydro production from peak to off-peak, implying that any action mitigating market power or promoting competition will increase peak hydro production relative to off peak.

In short, under the MPMA price cap, whether the monopolist is bidding at or above the price cap, it will shift its hydro production from off-peak to peak compared to unconstrained monopoly, implying that the water resource is now used in a more efficient way. Therefore we have our Proposition 3.

**Proposition 3** (hydro production) *Under the MPMA price cap, a monopoly generator would shift hydro production to the peak period, implying water is utilized more efficiently.*

### 3.5 Social Optimum

As usual, the welfare function is defined as the total of consumer and producer surplus in both periods. The welfare maximization problem can be written as follows (note that the rebate disappears from the aggregation):

$$\begin{aligned}
 \text{Max} \quad & \frac{B}{2}[(q_1^T + q_1^H)^2 + (q_2^T + \bar{q}^H - q_1^H)^2] + [A_1 - B(q_1^T + q_1^H)](q_1^T + q_1^H) \\
 & + [A_2 - B(q_2^T + \bar{q}^H - q_1^H)](q_2^T + \bar{q}^H - q_1^H) - 2a \\
 & - b(q_1^T + q_2^T) - \frac{c}{2}[(q_1^T)^2 + (q_2^T)^2]
 \end{aligned} \tag{3.39}$$

Solving this unconstrained problem results in the following global optimum:

$$q_{o1}^H = \frac{1}{2B}(A_1 - A_2 + B\bar{q}^H) \tag{3.40}$$

$$q_{o2}^H = \frac{1}{2B}(A_2 - A_1 + B\bar{q}^H) \tag{3.41}$$

$$q_{o1}^T = q_{o2}^T = \frac{1}{2(B+c)}(A_1 + A_2 - 2b - B\bar{q}^H) \tag{3.42}$$

$$P_{o1} = P_{o2} = \frac{1}{2(B+c)}(A_1c + A_2c + 2Bb - Bc\bar{q}^H) \tag{3.43}$$

Compared to the monopoly quantities (equation (3.6)-(3.9)), to achieve the welfare maximum, thermal output needs to be higher and hydro production in the off-peak (peak) must be lower (higher).

In a perfectly competitive context, the hydro allocation should go to where opportunity cost is highest, i.e. where the equilibrium price is the highest, until inter-temporal prices become equal. This is exactly the result of welfare maximization, since equation (3.43) shows the prices for both periods are equal. The equilibrium result is different from the

peak load pricing models (Steiner's Leontief production function model (1957) and Panzar's neo-classical production function model (1976)), which show the peak consumers should pay a higher price when supply cost is exogenous and fixed. The key factor that leads to the difference is that in the current model the production cost function is endogenous (resulting from a flexible hydroelectric resource).

Therefore, the socially optimal price cap should be:

$$P_{cap}^o = \frac{1}{2(B+c)}(A_1c + A_2c + 2Bb - Bc\bar{q}^H) \quad (3.44)$$

which is also equal to marginal cost of thermal production. The optimal price depends positively on the demand levels in both periods and the cost parameters, and is negatively related to the available water stock and the slope of the demand function.

Therefore when the policy is optimally designed, the price cap, marginal cost curve, and demand curve will intersect at the same point. As a consequence, the minimum  $CRQ$ 's are respectively:

$$Q_1^* = \frac{A_1(2B+c) - A_2c - 2Bb + Bc\bar{q}^H}{2B(B+c)} \quad (3.45)$$

$$Q_2^* = \frac{A_2(2B+c) - A_1c - 2Bb + Bc\bar{q}^H}{2B(B+c)} \quad (3.46)$$

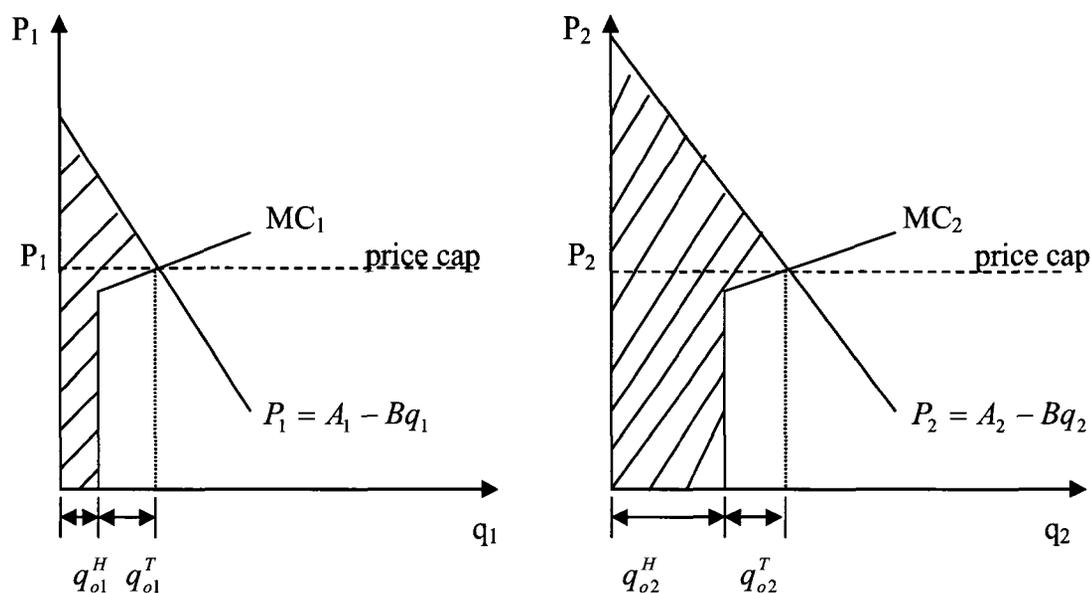
Of course, as demonstrated in *Example 3.2*, the optimal  $CRQ$ 's can be greater than (3.45) and (3.46). As long as the price cap is optimally set up and *the  $CRQ$ 's are at least as great as (3.45) and (3.46)*, the monopolist will bid the price cap. Although there are multiple optimal  $CRQ$ 's, we will define the  $CRQ$ 's determined by (3.45) and (3.46) as the

optimal  $CRQ$ s. Greater  $CRQ$ 's may not be accepted by the firm since those numbers are beyond the potential demand level. Thus, in the later analysis, we will assume the policy is socially optimally designed such that (3.44), (3.45) and (3.46) hold.

If the  $CRQ$ 's under the rebate mechanism are set to be 90% of the optimum production as in the case of the MPMA, we can immediately conclude that the bidding price must be higher than the price cap, as an equivalent result to Chapter 1. The intuition here is that lower  $CRQ$ 's will move down the marginal revenue curves, and lead the  $MR$  to intersect the  $MC$  at lower output levels than the social optimum.

An interesting finding here is that welfare maximization requires the market clearing prices in both periods to be identical, which is very different from the traditional peak load pricing models. The reason for this difference is that we do not have a capacity cost here and hydro generation can be shifted in a frictionless fashion between periods. This is equivalent to storing some output in the low demand period and selling it in the high demand period. During peak hours, the firm will produce more hydro energy to shave the high demand, which shifts the marginal cost curve of the thermal unit to the right and thus lowers the price in the peak period. The underlying implication of price equalization is that social welfare maximization requires the marginal benefit of electricity to equal the marginal cost of thermal production. This in turn implies that the marginal cost of thermal production must be equal in both periods. The welfare maximization equilibrium is illustrated in Figure 3.6. Both prices are equal to the price cap, i.e. the marginal cost of the thermal unit.

**Figure 3.6: Welfare Maximization Equilibrium**



Example 3.3: Assume the demand functions for off-peak and peak are  $P_1 = 80 - q_1$  and  $P_2 = 120 - q_2$ , respectively. The monopolist's cost function for thermal generation is  $C_i(q_i^T) = 3 + 2q_i^T + 0.1(q_i^T)^2$ . The maximum available water  $\bar{q}^H$  is 60MW. According to equations (3.44)-(3.46), the socially optimal price would be \$13.33/MWh, and the minimum optimum CRQ's are 66.67MW for off-peak and 106.67MW for peak. Because the CRQ's don't enter into the optimal production quantities and prices, any CRQ's that are greater than the minimum optimal CRQ's are socially optimal.

Under the socially optimal policy, the monopolist will produce 10MW of hydro power off-peak, 50MW on-peak, and 56.67MW of thermal electricity in each period. The profit under the optimal policy is \$1435.64, and social welfare is \$15036.58.

Now assume the CRQ's are set to be 90% of the optimal quantities, i.e. 60MW for off-peak and 96MW for peak. The regulated monopolist will bid above the price cap. This results in hydro outputs of 11MW and 49MW respectively, thermal output at 52.73MW for each period, based on bidding prices of \$16.27 and \$21 respectively. Profit will be \$2400.21 which is much greater than \$1435.64, the profit from bidding at the price cap.

**Proposition 4** (*first best equilibrium*) *the first-best solution lies on the “no rebate line”, and*

1. *The first best price cap is unique;*
2. *The first best CRQ's are multiple, as long as they are greater than or equal to the minimum optimal CRQs; and*
3. *The equilibrium price is the same in both periods.*

### **3.6 Impact of Changes in the Water Stock**

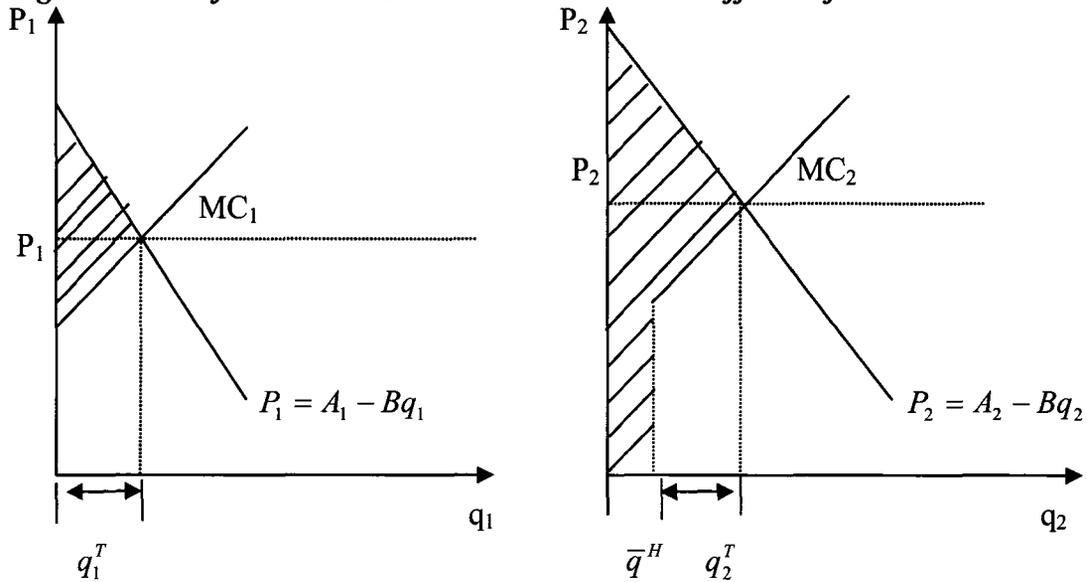
The above solution to the welfare maximization problem assumes that the water stock is large enough and the firm can profitably allocate some of its hydro production to each period. However, if the available water stock is so large that the thermal electricity production is not needed, the market price would be zero and the optimal price cap would be zero.

On the other hand, if the available water stock is so small that it is not profitable to allocate water to both periods, the entire hydro resource will be used in the peak period. In this case, off-peak has a lower thermal production and hence a lower price. At the same time, there is not enough water to offset peak thermal production to the same level. Figure 3.2 illustrates the equilibrium.

It's straightforward to derive the upper bound of water availability which drives both thermal production and prices to zero:

$$\bar{q}^H \leq \frac{A_1 + A_2 - 2b}{B} \quad (3.47)$$

**Figure 3.7: Welfare Maximization When Water is Insufficient for Both Periods**



The lower bound exists when the second period's price is exactly equal to the first period's price after the monopolist puts all hydro into second period production. By some manipulations, the lower bound can be written as

$$\bar{q}^H = \frac{A_2 - A_1}{B} \quad (3.48)$$

Substituting the welfare maximization equilibrium into the monopolist's profit function results in:

$$\begin{aligned} \pi_o = & -2a + \frac{1}{4(B+c)^2} [c(A_1 + A_2 - 2b)^2 + 2\bar{q}^H (A_1 c^2 + A_2 c^2 + 4Bbc + 2B^2 b) \\ & - Bc(B + 2c)(\bar{q}^H)^2] \end{aligned} \quad (3.49)$$

The profit could be positive or negative. If the fixed/start-up cost  $a$  is relatively high, the monopolist will experience a loss, which requires subsidies from the central planner. If a

subsidy is unavailable due to political constraints, a second-best Ramsey solution would be an alternative. On the other hand, if the fixed/start-up cost is relatively small, the monopolist can make a positive profit, and the first best solution is feasible.

When Ramsey pricing is required, the optimization problem can be expressed as a two-stage problem. In the first stage, the regulator chooses the price cap and *CRQs* that maximize consumer surplus, subject to zero profit to the firm. In the second stage, the firm maximizes its profit by choosing a quantity for each period, given the price cap and *CRQs*. A solution to the Ramsey problem is a price cap equal to average cost. Then it follows that, compared to the first best solution, thermal production under Ramsey pricing will be lower in both periods to allow a higher price (to cover the fixed cost), and hydro production will shift to the off-peak (because of lower shadow price of the water).

Taking the upper and lower bounds for the water stock into account gives the following corollary.

*Corollary 3.1* *Socially optimal prices are equal in both periods if and only*

$$\text{if } \frac{A_2 - A_1}{B} \leq \bar{q}^H \leq \frac{A_1 + A_2 - 2b}{B} .$$

### 3.7 The Impact of Demand Shocks

After the “optimal” policy is designed, there may be unexpected demand shocks resulting from either heat waves or cold waves, or from changes in economic activity. In a one-shot game, if a shock is unexpected, generation cannot respond to it and a price spike will

result. In a multi-period game, the firm can adjust its bidding strategies to respond to the shocks by changing its production schedule in later periods. Since analyzing the impact of a demand shock on Case 1 and Case 2 is not of particular interest to the policy-maker, our assessment will address Case 3 only. That is, the pre-shock policy is supposed to be socially optimal, and the monopolist is supposed to bid at the price cap if demand comes as expected. In other words, the equilibrium equations (3.44)-(3.46) hold.

To reflect the fact that a thermal generating unit has a much lower ramping rate than a hydro unit (i.e. a thermal unit cannot increase or decrease output as fast as a hydro unit), and thus takes much longer to react to a demand shock, we assume the thermal unit cannot respond to the demand change in the period in which the shock occurs, but can respond in the following period. In Scenario 1, the hydro unit is assumed not to react to the demand shock, while in Scenario 2 the hydro unit is able to respond to the shock. In both cases, we assume the shock is temporary and exists only for one period.

#### *Scenario 1 – Non responsive hydro*

There are two situations: the demand shock occurs either in the first period or the last period.

Let's first look at the situation when a positive demand shock happens in the last period.

It is straightforward to conclude that the firm still makes zero profit in the case of a positive shock. Because the monopolist cannot respond to the shock by adjusting its

production at either the thermal or the hydro station, because there is no additional water available in the last period, the market clearing price in the second period must be higher than the pre-shock equilibrium, therefore the average price must be higher than the price cap. Consequently, the monopolist has to refund some revenue to consumers. The net change in profit will be zero since the post-shock profit  $B\varepsilon(q_{o2} - Q_2^*)$  is equal to zero by the assumption of  $q_{o2} = Q_2^*$ , where  $\varepsilon$  is the amount of change in the demand (either positive or negative),

An interesting result in this case is that the first period consumers are better off.

Although the price in the first period doesn't change as a result of the demand shock in the second period, the consumers still receive refunds in proportion to their total consumption. On the other hand, the consumption in the second period is rationed because production cannot be increased in response to the demand shock.

In contrast, if the demand shock is negative, the last period price will be lower than the price cap, and consequently the firm will make less revenue. Given that the policy allows the monopolist a zero profit for the expected higher demand condition, the monopolist now suffers a loss (i.e. the area representing producer surplus is less than the fixed cost that has been incurred to produce energy).

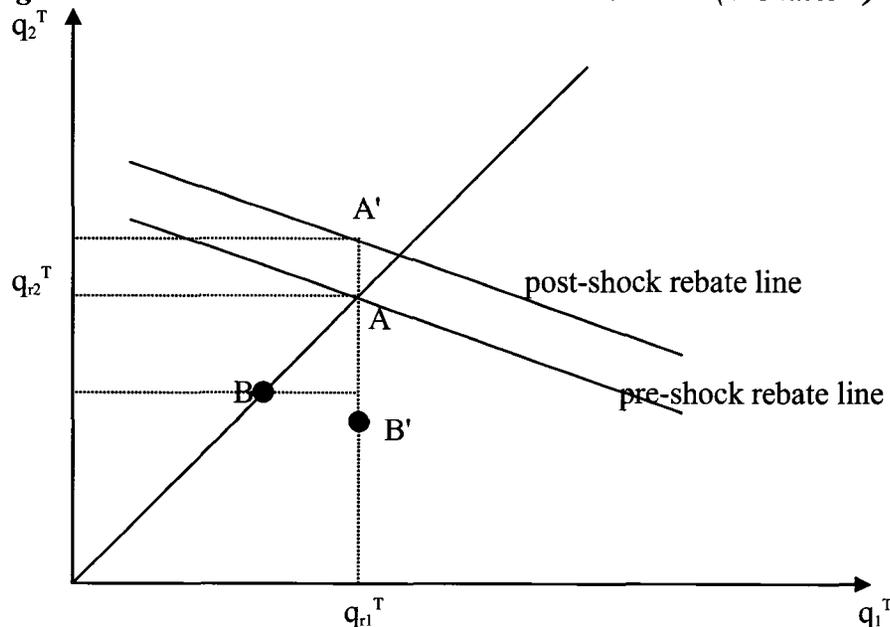
For the above reasons, in the existence of a demand shock, the firm bears all risks.

Therefore, a lower  $CRQ$  than the expected social optimum (e.g. 90% of optimal output as stated in the MPMA) could have the effect of risk sharing between the monopolist and

consumers. In good or normal states, the firm is allowed to make a positive profit; while in bad states, it can break even or suffer a small loss.

If the demand shock happens in the first period, in contrast, the monopolist could adjust its thermal production in the second period, and charge a different price that maximizes its aggregate profit. Because the pre-shock equilibrium is of Case 3, the post-shock situation cannot be of Case 1.

**Figure 3.8: Thermal Production with a Demand Shock (Scenario 1)**



Prior to the demand shock, the monopolist will produce at point A, where its profit is the highest among all strategies, and equals zero so as to maximize the social welfare. The demand shock moves the rebate line up, and the monopolist could produce at point A' (on the new rebate line), or at point B' (above the price cap). Therefore the second period's post-shock thermal production may increase or decrease, depending on the monopolist's strategy.

First, assume the demand shock is positive. If the monopolist chooses to bid above the price cap, its optimal thermal output in the second period would be  $q_{b2}^{T'} = \frac{A_2 - b + BQ_2^* - 2Bq_{o2}^H}{2B+c}$ ,

which is smaller than the pre-shock thermal output. The new production mix is point  $B'$  in Figure 3.8. However, the firm won't produce at that point because by doing so it will make a loss, which is even worse than producing at the scheduled level to make zero profit. Because the monopolist cannot change its strategy in the first period, the price will go up by  $B\varepsilon$ . If the firm continues to produce  $q_{o2}^T$  in the second period, it will still make zero profit after rebate. In contrast, if it bids a high price in the second period by withholding its capacity, it will have to pay additional “penalty” for production below  $CRQ$  (the “penalty”, i.e. refund, could be avoided if the firm produced at  $CRQ$ . See the section on outages in Chapter 2 for more details).

However, the monopolist still has a third strategy: increase peak thermal production so that the average price is equal to the price cap (point  $A'$ ). The monopolist will do so if the cost increase (for a higher thermal output) minus the revenue decrease in the second period (for a lower price resulting from a higher thermal output) is less than the revenue increase from the first period production. It can be shown that the post-shock profit of bidding at the price cap is:

$$\pi' = -\frac{(2B + c)\beta^2\varepsilon^2}{2} \quad (3.50)$$

which is negative. Therefore the monopolist will never increase its production to stick to the price cap.

Thus, in case of a positive demand shock, the monopolist will continue its pre-shock strategy and make zero profit. Consumers in the later period, however, are better off from receiving refund due to a high price paid by the consumers in the first period.

If the demand shock is negative, the firm will not bid above the price cap because by doing so the monopolist will earn lower revenue in the off-peak period, while paying a “penalty” for output lower than  $CRQ$  in the peak period. Therefore in the case of a negative shock, there is only one possibility: bidding at the price cap<sup>27</sup>. This is a similar result to the one period model. Although bidding at the cap means that the firm is making a loss, this strategy is its best choice among all alternatives. Thus, in equilibrium, first period consumers see a drop in price, while second period consumers pay a higher price.

In contrast, for an unconstrained monopolist, the demand shock (positive or negative), either in the first period or in the second period, has no impact on its supply schedule. The key point here is that there is no interdependence between the two periods: no hydro reallocation by assumption and no refund. Therefore, the monopolist will still follow the pre-shock schedule.

***Proposition 5 (demand shock 1: non-responsive hydro).***

1. *If the demand shock appears only in the last period, the monopolist bears all the risk. In particular, when the shock is positive, the bidding price will be higher than the price cap but the firm still makes zero profit. In this case, the first period consumers are better off; if the shock is negative, however, the bidding price will be lower than the price cap and the monopolist suffers a loss.*

---

<sup>27</sup> Bidding below the price cap is not rational because the monopolist can be actually better off by withholding some capacity in the peak period to increase its profit.

2. *If the demand shock happens in the first period, the monopolist can share some risk with consumers by charging a higher price to second period consumers.*

*Specifically, in the event of a positive shock, the firm will continue to produce its pre-shock optimum output (resulting in a higher price in the first period) and makes zero profit. In this case, consumers in the second period are better off. If the demand shock is negative, the monopolist will reduce its thermal production in the second period and bid at the price cap, but makes negative profit. Then consumers in the second period also pay the price.*

#### *Scenario 2 -- Responsive hydro*

While most fossil fuelled generators cannot quickly respond to exogenous shocks, hydro generators have a much higher ramp rate and thus can quickly smooth the shocks. In the electricity industry, the ramp rate of a typical fossil fuel generating unit is about 5~10Mw per minute, while a hydro unit can within a few minutes ramp up from zero to its full capacity or down from full capacity to zero. Thus hydro generation can in theory smooth almost any demand fluctuation if there are sufficient hydro resources in a system.

Again we will only look at the shock that occurs in the first period. If the demand shock is positive, the firm has three choices as before: continuing its pre-shock schedules (implying a higher price), changing its production schedule while bidding above the price cap, or changing its schedule and bidding at the price cap. The first strategy will obviously result in zero profit, because all extra revenue from higher demand goes back

to consumers. The second strategy will lead to more hydro production in the first period to shave the suddenly increased demand, and more thermal output in the second period to offset the lower hydro electricity production. The third strategy requires the monopolist to increase second period thermal production sufficiently to offset the impact of a higher first period price.

If the firm changes its production mix and bids above the price cap, the equilibrium outcome would be:

$$q_{b1}^{T'} = q_{o1}^T \quad (3.51)$$

$$q_{b2}^{T'} = q_{o2}^T + \frac{B\varepsilon}{2(B+c)} \quad (3.52)$$

$$q_{b1}^{H'} = q_{o1}^H + \frac{\varepsilon(2B+c)}{4(B+c)} \quad (3.53)$$

$$q_{b2}^{H'} = q_{o2}^H - \frac{\varepsilon(2B+c)}{4(B+c)} \quad (3.54)$$

Water is now shifted from the second period to the first period. The aggregate output in the first period ( $q_{b2}^{H'} + q_{b2}^{T'}$ ) is equal to  $q_{o2}^T + q_{o2}^H - \frac{c\varepsilon}{4(B+c)}$ , which is less than the pre-shock optimum  $q_{o2}^T + q_{o2}^H$ . Because the pre-shock second period optimal  $CRQ$  is  $q_{o2}^T + q_{o2}^H$ , the firm is now producing below its  $CRQ$ .

The profit of this strategy is:

$$\pi_b' = B\varepsilon(q_{o1}^H + q_{o1}^T) - \frac{B\varepsilon^2(2B+c)}{8(B+c)} \quad (3.55)$$

which is greater than zero because the first term will be generally greater than the second one unless the shock is extremely large. As a result, the monopolist can be better off by

bidding above the price cap. Because second period production is now lower, the price must be higher, and thus the transitory shock here has a permanent price impact.

If the monopolist is still following the strategy of bidding at the price cap, the optimal production would be as follows:

$$q_{r1}^{H'} = q_{o1}^H + \frac{\varepsilon[B+2B\beta^2-c\beta(1-\beta)]}{2B+2B\beta^2+c(1-\beta)^2} \quad (3.56)$$

$$q_{r2}^{H'} = q_{o2}^H - \frac{\varepsilon[B+2B\beta^2-c\beta(1-\beta)]}{2B+2B\beta^2+c(1-\beta)^2} \quad (3.57)$$

$$q_{r1}^{T'} = q_{o1}^T \quad (3.58)$$

$$q_{r2}^{T'} = q_{o2}^T + \frac{B\varepsilon(1+\beta+2\beta^2)}{2B+2B\beta^2+c(1-\beta)^2} \quad (3.59)$$

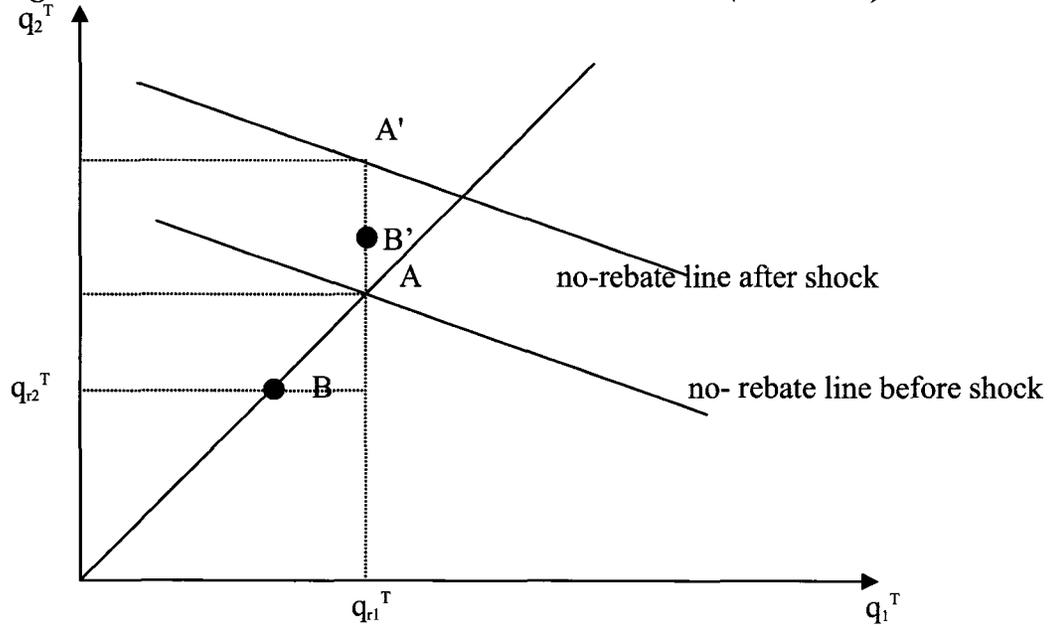
One can see that the hydro production is shifted to the first period, and there is more thermal electricity supply in the second period. The result is intuitive: Once the positive shock comes along, the off-peak price must increase unless production is increased. To lower the price, the firm has to shift its hydro production to smooth the shock, which in turn leaves less water for the second period, leading to a higher peak price if its thermal output is not increased. Therefore, to keep the average price at the price cap level, the monopolist has to increase its hydro output in first period and its thermal output in the second period.

It is straightforward that  $q_{r2}^{T'} > q_{r1}^{T'}$  because  $q_r > q_b$ . Thus  $\frac{B\varepsilon(1+\beta+2\beta^2)}{2B+2B\beta^2+c(1-\beta)^2} > \frac{B\varepsilon}{2(B+c)}$ .

Figure 3.9 illustrates the above results. If the monopolist chooses to bid above the cap and adjusts its output level, its second period thermal production will be higher (point  $B'$ )

than the scheduled output (point  $A$ ). If the firm produces at the price cap, its second period thermal output (point  $A'$ ) has to be even higher than  $B'$ .

**Figure 3.9: Thermal Production with a Demand Shock (scenario 2)**



Substituting equations (3.56)-(3.59) into the profit function of bidding at the cap yields:

$$\pi_r' = Bw(1-w)\varepsilon^2 - B\beta^2(1-w)^2\varepsilon^2 - \frac{c}{2}(w + \beta - \beta w)^2\varepsilon^2 \quad (3.60)$$

where  $w = \frac{B+2B\beta^2-c\beta(1-\beta)}{2B+2B\beta^2+c(1-\beta)^2}$ .  $W$  is greater than  $\frac{1}{2}$  but less than 1, because  $c$  is much smaller

than  $B$  (i.e. the marginal cost curve has a smaller slope than the demand function by

assumption) and  $0 < \beta < 1$ . Therefore,  $\pi_r'$  must be greater than zero. That is, the

monopolist can be better off by changing its production schedule and bidding at the price

cap.

Although it is not straightforward comparing  $\pi'_r$  and  $\pi'_b$ , it is possible to show that  $\pi'_b$  must be larger than  $\pi'_r$ . From equation (3.55), the profit  $\pi'_b$  is dominated by the first term  $B\varepsilon(q_{o1}^H + q_{o1}^T)$ . As long as the shock  $\varepsilon$  is relatively small, the positive term in  $\pi'_r$ ,  $Bw\varepsilon^2(1-w)$ , is far smaller than  $B\varepsilon(q_{o1}^H + q_{o1}^T)$ . In contrast, the aggregation of the last two terms in  $\pi'_r$  cannot be too different from the second term in  $\pi'_b$ . For this reason  $\pi'_b$  will generally be greater than  $\pi'_r$ . This result resembles the situation in which hydro is not responsive.

If the demand shock is negative, following the same logic of scenario 1 in which hydro is not responsive, the strategies of bidding above or below the price cap can be excluded, and thus the only solution is to bid at the price cap. In this situation, the monopolist will shift its hydro to the second period. Its hydro electricity output in period 1 will be  $q_{o1}^H - \varepsilon$ . Given no changes except for more supply of hydro electricity, the equilibrium price in the second period must be the price cap (at the disjoint point of MR curve), following the analysis in Chapter 2. As a result, the equilibrium prices in both periods need to be equal to the price cap, and aggregate output is reduced by  $\varepsilon$ .

Similar to the one period model with a negative demand shock, the monopolist will make a loss. The loss results from reduced thermal output which has an upward sloping marginal cost curve. The lost profit is a right triangle with an area equal to  $\frac{1}{2}c\varepsilon^2$ . The smaller  $c$  is, the smaller loss is.

In summary, when hydro production can shift inter-temporally, the second period consumer will bear some of the risk of a positive demand shock that happens in first period and pay a higher price, and the monopolist will make a positive profit. This is contrast to the case where the monopolist has no hydro resource. In that case, the monopolist will continue its old strategy and make a zero profit. When the demand shock is negative, the first period consumers won't see a price drop, and the monopolist makes a loss.

Interestingly, a regulated monopolist will respond to a positive demand shock no more efficiently than an unregulated monopolist. To see this, solving the unconstrained monopoly problem yields:

$$q_{m1}^{H'} = q_{m1}^H + \frac{\varepsilon(2B+c)}{4(B+c)} \quad (3.61)$$

$$q_{m2}^{H'} = q_{m2}^H - \frac{\varepsilon(2B+c)}{4(B+c)} \quad (3.62)$$

$$q_{m1}^{T'} = q_{m1}^T \quad (3.63)$$

$$q_{m2}^{T'} = q_{m2}^T + \frac{B\varepsilon}{2(B+c)} \quad (3.64)$$

Facing a positive demand shock, an unconstrained monopolist will shift exactly the same amount of hydro from second period to first period as the constrained monopolist (note the second term in (3.61) is the same as the last term in (3.53)), and thus increase the same amount of thermal production in the second period. Thus the equilibrium price in both periods under unconstrained monopoly will increase by same amount as does the constrained monopolist. In this sense, the unconstrained monopoly and constrained monopoly are both efficiently responding to the demand shock.

Facing a negative demand shock, however, an unconstrained monopolist will act more efficiently. The point is that in the absence of the MPMA, the firm will lower its aggregate production by an amount smaller than the shock (because of its downward-sloping MR), while the constrained monopoly will lower production exactly by the amount of the shock (because of the discontinuity of its MR).

A standard policy implication under demand uncertainty is that the “optimal” design of MPMA determined for the certainty case is not at all optimal when the regulated firm is facing uncertainty and/or has endogenous production costs. In particular, the optimal price cap under the rebate mechanism serves as a price floor: the equilibrium price will increase when demand increases, but won’t decrease as demand drops. Because of the lack of a simultaneous response of thermal production to the demand shock, the outcome will never be optimal. An analysis incorporating the probability distribution of shocks would be very complicated because the two-period model will necessarily involve an interaction between the probability distribution of uncertainty in both periods.

When demand uncertainty is the only factor affecting the regulator’s decision, the optimal policy under demand uncertainty might be quite simple as is shown in Lewis and Sappington (1988). They show that when marginal production cost is common knowledge and increases with output, the regulator can set the price at the marginal cost where demand is realized. However, when marginal cost is decreasing with output, the regulator can only rely on his/her imperfect information on demand to set a fixed price and allow the firm to extract some rent if any since it is too costly to monitor.

The existence of a flexible hydro resource also makes policy design more complicated because the shifting of water between periods has essentially changed the firm's cost function. That is, the firm's cost is now a function of the distribution of demand over time and thus varies with the firm's knowledge of demand. Given the discontinuity of the marginal revenue curve, any attempt to derive an optimal policy under demand uncertainty and endogenous cost would be extremely challenging.

**Proposition 6** (*demand shock 2: responsive hydro*) *In the case of a positive demand shock*

1. *If the demand shock appears only in the second period, the consequence will be the same as that in Scenario 1.*
2. *If a positive demand shock happens in the first period, the average price will be higher than the price cap. The second period consumer will share the cost of the demand shock in the first period. In this situation, the monopolist makes a positive profit (in contrast to a zero profit in the case of non-responsive hydro). In contrast, in the event of a negative demand shock, the equilibrium price would remain the same for each period, and the monopolist makes a loss.*
3. *If a positive demand shock happens in the first period, an unconstrained monopolist will respond to the demand shock in the same way as the constrained monopoly [i.e. increasing production by the same amount]. In contrast, if a negative shock occurs, an unconstrained monopolist will respond to the shock more efficiently.*

### 3.8 Imports

In Chapter 2, we found that the monopolist will import electricity to meet domestic demand when there is a supply outage and when the import price is not too high. In a two-period model, the firm has flexibility to shift its water resource or adjust its thermal production in the later period. For instance, if a forced outage occurs in the first period, the monopolist can import to meet domestic demand, or produce more in the second period, so as to stick to the price cap (Case 3), or charge a higher price than the cap and rebate some revenue to costumers (Case 2), whichever is more profitable.

Because a forced outage usually happens to fossil fuelled plants due to technical difficulties, let's assume that the outage happens only on the thermal unit. We further assume the monopolist cannot respond to this outage by changing its hydro production, but can import electricity in the period in which the outage happens, and can adjust its thermal production in the following period. In later sections we will assume the hydro unit can respond immediately to the supply shock.

Let  $q_a^T$  be the maximum available thermal capacity after the outage and  $P_{imp}$  the import price for both periods. Assume  $P_{imp}$  to be greater than the domestic price and production cost at any output level. Furthermore, assume that the outage happens only in one period.

First let's look at the situation where the outage happens in the first period. If importing is impossible, e.g. due to transmission congestion or an outage, the firm will have only

two choices in the second period: to produce more thermal energy and keep the average price down to the price cap, or simply take no action allowing the average price to shoot up. Because the outage drives the firm to deviate from its optimum, where it makes zero profit, it will now experience a post-outage loss. Therefore if an outage happens, the firm will adjust its strategy to minimize its total loss.

If the monopolist chooses to offer above the price cap, it can make one of two choices in the second period: continuing the scheduled production, or rescheduling it. Solving the problem assuming bidding above the price cap yields a lower peak thermal production than the pre-outage optimum, implying the monopolist must pay an additional refund for what it has never earned (the missed quantity between the  $CRQ$  and its actual production). Therefore the monopolist will make a smaller loss if it follows its original production strategy. Under this production strategy, the firm's profit is:

$$\pi'_b = P_{cap}^o (q_a^T - q_{o1}^T) - B(q_a^T - q_{o1}^T)^2 \quad (3.61)$$

which is negative as expected.

If the monopolist bids at the price cap by increasing the output in the second period, however, its profit will be:

$$\pi'_r = -\frac{(2B+c)(1+\beta^2)}{2} (q_a^T - q_{o1}^T)^2 \quad (3.62)$$

$\pi'_r$  is greater than  $\pi'_b$  if  $q_a^T > \bar{q}_a^1$ , where  $\bar{q}_a^1 = q_{o1}^T - \frac{2P_{cap}^o}{c(1+\beta^2)}$ . The condition can also be

rearranged as  $q_{o1}^T - q_a^T < \frac{2P_{cap}^o}{c(1+\beta^2)}$ . We define  $q_{o1}^T - q_a^T$  as the outage amount. Therefore, the

monopolist will bid at the price cap if the outage is less than  $\frac{2P_{cap}^0}{c(1+\beta^2)}$ . Although the average price is still equal to the price cap, the prices in the two periods diverge: a higher first period price and a lower second period price, in contrast to equal prices when there is no outage. Note that if the firm's marginal cost is close to constant, i.e. if  $c$  is close to zero, the above inequality is always satisfied. That is, bidding at the price cap is always the best choice if the marginal cost is nearly constant.

In summary, when importing is not allowed, the monopolist's decision rule is the following: bid at the price cap if the outage is relatively small, otherwise above it by producing at the originally scheduled level.

Now let's turn to the case where importing is allowed. In this situation, the monopolist may import electricity when the outage happens, if it is profitable to do so. Because the import price is assumed to be sufficiently high, the firm has no incentive to import in the last period no matter whether or not there is an outage in that period.

In the current setting, the monopolist has three choices: not importing, importing but bidding above the cap, or importing and bidding the cap. If the firm chooses not to import, the decision rule is the same as for the case where importing is not feasible: bid the price cap if the outage is less than  $\frac{2P_{cap}^0}{c(1+\beta^2)}$ , otherwise bid above the price cap.

If the monopolist imports power and reschedules its peak thermal but sells above the price cap, the monopolist's import and peak thermal output are respectively;

$$q_{b,imp} = q_{o1}^T - q_{a,1}^T + \frac{1}{2B}(P_{cap}^o - P_{imp}) \quad (3.63)$$

$$q_{b2}^{T''} = \frac{1}{2B+c}(A_2 - b + BQ_2^*) - q_{o2}^H \quad (3.64)$$

Since  $q_{b,imp} > 0$ ,  $P_{imp}$  must be less than  $P_{cap}^o + 2B(q_{o1}^T - q_{a,1}^T)$ . Thus, the larger the outage is, the more profitable importing will be. According to (3.64) and (3.42),  $q_{b2}^{T''} < q_{o2}^T$ .

Because its production is now lower than the  $CRQ$ , the firm has to pay a rebate for what it didn't produce. Thus, the monopolist can be better off by not changing the peak thermal schedule. This would be exactly the outcome in the static case.

If the monopolist imports energy and continues to sell power at the price cap, the quantity imported and thermal production will be respectively:

$$q_{r,imp} = \frac{1}{2B(1+\beta^2)+c\beta^2} [A_1 - A_2\beta + b\beta + 2B\alpha\beta - P_{imp} + c\beta(\alpha - \bar{q}^H) + c\beta q_{o1}^H] - (q_{r1}^H + q_{a,1}^T) \quad (3.65)$$

$$q_{r2}^{T''} = \frac{1}{2B(1+\beta^2)+c\beta^2} [-\beta(A_1 - A_2\beta + b\beta + 2B\alpha\beta - P_{imp}) + 2B(1+\beta^2)(\alpha - \bar{q}^H + q_{o1}^H)] \quad (3.66)$$

The import (equation (3.65)) can be rearranged as:

$$q_{r,imp} = \frac{q_{o1}^H(1-\beta)(2B+2B\beta^2+c+c\beta)+(\alpha-\bar{q}^H)(2B+2B\beta^2+c+c\beta^2)}{(1-\beta)(2B+2B\beta^2+c+c\beta^2)} - q_{a,1}^T - \frac{(P_{imp}-b)}{2B+2B\beta^2+c\beta^2} \quad (3.67)$$

Because the first term is approximately equal to  $q_{o1}^T$ ,  $q_{r,imp}$  then

approximates  $q_{o1}^T - q_{a,1}^T - \frac{(P_{imp}-b)}{2B+2B\beta^2+c\beta^2}$ , i.e. the *outage adjusted for the import price*. For

simplicity of manipulation, we take this approximation as the quantity imported.

Substituting the equilibrium into the respective profit functions yields the differential between the profit of bidding the cap and the profit of bidding above the cap:

$$\Delta\pi = \frac{(P_{imp}-b)^2}{4B+4B\beta^2+2c\beta^2} - \frac{(P_{imp}-b)(A_1-A_2\beta-b+b\beta+c\beta q_{o2}^T)}{2B+2B\beta^2+c\beta^2} - \frac{(P_{cap}^O-P_{imp})^2}{4B} - \frac{(P_{imp}-P_{cap}^O)(q_{o1}^T+q_{o1}^H)}{2} \quad (3.68)$$

$\Delta\pi$  is negative, implying that the monopolist will never import electricity and offer at the price cap at the same time. This result is consistent with the result in the one period model, showing that the monopolist will offer above the price cap whenever there is a large outage.

In summary, if a forced outage occurs in the first period and importing is allowed, the monopolist's decision rule is:

1. if  $q_{a,1}^T > q_{o1}^T - \frac{P_{imp}-P_{cap}^O}{2B}$ , follow the rule where no importing is assumed; or
2. if  $q_{a,1}^T < q_{o1}^T - \frac{P_{imp}-P_{cap}^O}{2B}$ , import power and bid above the price cap in the period when the outage happens.

If the forced outage occurs in the second period, the situation resembles the single period case (in Chapter 2), where the monopolist will import if the import price is not too high. However, the situation with two periods becomes more complicated because a higher price in one period will lead to a higher average price, which in turn will affect its revenue in another period. Therefore, the firm has stronger incentive to lower the price in the outage period by importing electricity. From the rebate formula, we know that the peak price has a higher weight, and thus the firm has stronger incentives to mitigate the impact on the peak price.

Because of lack of interdependence between two periods on both cost and demand side, a monopolist facing a forced outage in either period and not subject to the MPMA will act as if it maximizing profits in two separate markets. The decision rule for importing will simply be the same as in the static case. Specifically, an unconstrained monopolist won't import any electricity if the available capacity is greater than  $\frac{A_i - P_{imp}}{2B}$ ,  $\forall i = 1, 2$ , and will import otherwise.

**Proposition 6** (*forced outage 1: inflexible hydro production*)

*I. If the outage happens in the (off-peak) first period and importing is impossible, e.g. because of intertie congestion, the firm's decision rule will be as follows:*

- 1. bid at the price cap if the outage  $q_{o1}^T - q_a^T < \frac{2P_{cap}^o}{c(1+\beta^2)}$ ; or,*
- 2. bid above the price cap with no change in the second period production, otherwise.*

*II. If importing is possible, the monopolist's decision rule will be:*

- 3. if the outage  $q_{a,1}^T - q_{o1}^T < \frac{P_{imp} - P_{cap}^o}{2B}$ , follow rule (I); or*
- 4. Otherwise, import power and bid above the price cap in the period when the outage happens.*

*III. If the forced outage happens in the second period and the*

*outage  $q_{o2}^T - q_{a,2}^T > \frac{P_{imp} - P_{cap}^o}{2B}$ , the monopolist will import electricity but bid above the*

*price cap. Otherwise it will not import but bid above the cap (as in the single period case).*

Now, let's turn to the case when the hydro production is flexible, i.e. hydroelectric power can quickly respond to the forced outage. We will only look at the case when the outage occurs in the first period, when the monopolist can shift its water resource from the second period to respond to the supply shock.

When importing is not possible, the monopolist has only two choices: bidding above or at the price cap. If it bids above the price cap, its first-period optimal hydro output and second-period optimal thermal output are respectively:

$$q_1^H = q_{o1}^H + \frac{2B+c}{2(B+c)}(q_{o1}^T - q_{a1}^T) \quad (3.69)$$

$$q_2^T = q_{o2}^T + \frac{B}{B+c}(q_{o1}^T - q_{a1}^T) \quad (3.70)$$

That is, the monopolist will shift hydro production to the outage period and increase thermal output in the following period. As a result, its profit would be:

$$\pi_b = - \frac{(2B+c)c}{2(B+c)}(q_{o1}^T - q_{a1}^T)^2 \quad (3.71)$$

which is obviously negative, and negatively related to the amount of the outage.

In contrast, if the monopolist chooses to bid at the price cap, the resulting off-peak hydro production and peak thermal production are respectively:

$$q_{r1}^H = q_{o1}^H + \frac{2B+2B\beta^2-c\beta(1-\beta)}{2B+2B\beta^2+c(1-\beta)^2}(q_{o1}^T - q_{a1}^T) \quad (3.72)$$

$$q_{r2}^T = q_{o2}^T + \frac{2B+2B\beta^2}{2B+2B\beta^2+c(1-\beta)^2}(q_{o1}^T - q_{a1}^T) \quad (3.73)$$

Apparently, both quantities are increased given that the outage  $q_{o1}^T - q_{a1}^T$  is positive. In this case,  $q_{a1}^T + q_1^H$ , is lower than the original optimum, while the aggregate output in period 2,  $q_2^T + q_2^H$ , is greater than the original optimum. The corresponding profit is thus:

$$\pi_r = -\frac{c[4B+4B\beta^2+c(1-\beta)^2]}{2[2B+2B\beta^2+c(1-\beta)^2]} [q_{o1}^T - q_{a1}^T]^2 \quad (3.74)$$

which is also negative.

Therefore,  $\pi_b > \pi_r$ , if  $2 + 2\beta^2 > (1 - \beta)^2$ . Given that  $0 < \beta < 1$ , the condition is always satisfied. That is, the monopolist will always bid above the price cap if importing is not possible and hydro shifting is allowed.

Will the monopolist import if importing is permitted? The answer is no. Because the use of water is flexible, the monopolist can use as much water as it wants in the first period when the outage happens, while producing more thermal electricity in the following period to offset the reduction in hydro production. By assumption, the import price is greater than the domestic production cost at any production level for the thermal unit. Thus the monopolist has no incentive to buy imported electricity when it faces a forced outage situation.

In summary, when the monopolist can adjust its hydro production, it won't import any electricity following a forced outage in the first period. Moreover, the monopolist will

always bid above the price cap, shift hydro production to the first period, and increase thermal production in the second period.

***Proposition 7 (forced outage 2: flexible hydro production)***

*If the outage happens in the first period, the monopolist will not import any electricity, always bid above the price cap, shift hydro production from the second to the first period, and increase thermal production in the second period.*

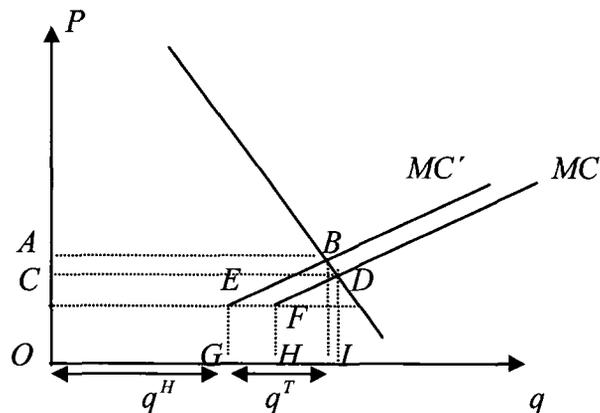
### **3.9 Marginal Cost Pricing**

It is argued that if the first best policy is marginal cost pricing, why not simply require the monopolist to bid marginal cost and subsidize it whenever there is a deficit. It is true that marginal cost pricing is the first best in a static single-plant model when the cost function is known and the output is observable. In a multi-period and multi-plant model like the present case, however, marginal cost pricing may not be socially optimal. The intuition is that the social optimum requires equal prices in both periods, which will not be to the best interest of the monopolist. For example, the monopolist may shift its hydro production from one period to another, or spill water, so long as the increased revenue resulting from higher price in one or both periods is greater than the increased cost resulting from higher output of thermal production.

Figure 3.10 illustrates the possibility that the monopolist could be better off by spilling water. Assume the socially optimal point is at  $D$ , where the monopolist will have to

produce  $OH$  units of hydro electricity and  $HI$  units of thermal electricity. However, the monopolist may be better off if the gain in revenue  $ABDCA$  (approximate) resulting from spilling  $HG$  units of hydro power is greater than the cost  $BDFHGEB$ . The same is true for the other period. Consequently, the spilled water need not go to the other period. This is obviously inefficient because low cost hydro is wasted. This illustration highlights the importance of monitoring the usage of hydro resources even though the marginal cost of thermal production is easily verified.

**Figure 3.10: Profitability of Spilling Water**



One way to prevent the spilling of valuable water is to auction the water resources (for example, selling those hydro stations to various competitors, following Demsetz's "competition for the market") or to regulate it through a contract for difference approach. The monopolist may win the auctions but it would not spill water any more because it has paid a competitive price for it. However, this is still unable to prevent the monopolist from using the water resource inefficiently because of its large portfolio.

To see how the monopolist reacts to the marginal cost pricing rule (assume no water spill), rewrite its objective function as follows:

$$\begin{aligned}
 \text{Max}_{q_i^H, q_i^T} : & \sum_{i=1,2} (P_i(q_i^H + q_i^T) - C(q_i^T)) & (3.75) \\
 \text{s.t.} & (1) P_i = A_i - B(q_i^H + q_i^T) \quad \forall i = 1,2 \\
 & (2) C(q_i^T) = a + bq_i^T + \frac{1}{2}c(q_i^T)^2 \\
 & (3) \sum_{i=1,2} q_i^H = \bar{q}^H \\
 & (4) P_i = MC(q_i^T)
 \end{aligned}$$

Because the price in each period must be equal the marginal cost of thermal electricity in that period, the constraint (1) and (4) can be combined to solve the supply of  $q_1^T$  and  $q_2^T$  as function of  $q_1^H$ . That is, the objective function can be expressed as a function of  $q_1^H$ .

Solving the maximization problem yields the following equilibrium:

$$q_1^H = \frac{(A_1 - A_2)c}{2B(B+2c)} + \frac{1}{2}\bar{q}^H \quad (3.76a)$$

$$q_2^H = \frac{(A_2 - A_1)c}{2B(B+2c)} + \frac{1}{2}\bar{q}^H \quad (3.76b)$$

$$q_1^T = \frac{1}{2(B+c)} \left[ \frac{(2B+3c)A_1 + A_2c}{B+2c} - 2b - B\bar{q}^H \right] \quad (3.76c)$$

$$q_2^T = \frac{1}{2(B+c)} \left[ \frac{(2B+3c)A_2 + A_1c}{B+2c} - 2b - B\bar{q}^H \right] \quad (3.76d)$$

$$P_1 = \frac{1}{2(B+c)} \left[ \frac{A_1c(2B+3c) + A_2c^2}{B+2c} + 2Bb - Bc\bar{q}^H \right] \quad (3.76e)$$

$$P_2 = \frac{1}{2(B+c)} \left[ \frac{A_2c(2B+3c) + A_1c^2}{B+2c} + 2Bb - Bc\bar{q}^H \right] \quad (3.76f)$$

One way to see the outcome is not the social optimum is that the prices are not equal, given that  $A_2 > A_1$  by assumption. It is of interest to note that the total thermal production  $q_1^T + q_2^T$  is exactly the same as the total output under the social optimum. That is, the marginal cost pricing monopolist will produce the same amount of electricity as the social optimum, but less on-peak and more off-peak.

This problem is a moral hazard problem and the regulator is unable to induce the first best solution in a monopoly framework (Laffont and Tirole, 1986, 1994). It is, however, possible to achieve the social optimum if the monopolist is split—leave one firm with thermal capacity only who is subject to a marginal cost pricing requirement and sell hydro resources to as many small owners as possible. The small hydro firms, who would have little market power and would thus be price takers, would allocate their water resource so that marginal revenue in each period is equal, i.e. until the peak and off-peak prices are the same<sup>28</sup>. On the other hand, the thermal firm will generate power along its marginal cost curve, as required, and produce the same amount of power in each period because the price is the same in each period, as a result of competitive hydro generation. Therefore, marginal cost pricing by a thermal monopolist could be socially optimal if hydro generators behaved competitively.

Thus we have the following proposition.

***Proposition 8*** *As an alternative to the MPMA, marginal cost pricing generally won't lead to a social optimum; the water resource will be inefficiently used. To achieve the social optimum through marginal cost pricing, the monopolist could be dissolved into one firm with thermal capacity only and subject to marginal cost pricing, and into either price-taking hydro firms, or a hydro monopolist subject to a contract for difference procurement arrangement.*

---

<sup>28</sup> If the hydro facility has a capacity limitation, it would produce at its maximum capacity when the price is highest, and shut down, if possible, when the price is lowest.

### 3.10 Conclusions

The hydro reserve plays a key role in the above two-period model. The model is equivalent to a standard inventory model in the peak load pricing literatures, e.g. Gravelle (1976), where the output can be produced off-peak and sold on-peak. In our case, output—electricity—is non-storable, but the input—the water stock—is storable and can be converted into output without cost.

However, there is a fundamental difference between our model and traditional peak load models with inventory: the flexibility of hydro production changes the aggregate supply relation, i.e., the generation cost function, while in the inventory model it doesn't. It is this difference that leads to no excess capacity requirement for the traditional peakers (usually thermal units) in our peak-load context. Instead, the excess capacity of hydro units (because some hydro capacity will be idle in the low demand periods) plays a critical role in shaving the peak demand. As a consequence, it is possible that some investment in thermal capacity could be avoided if the hydro resource could be better utilized. That is, a well functioning regulation that allocates the hydro resource in the most efficient way could effectively avoid some unnecessary investments in peaking fossil generation.

The analysis in this chapter has produced a number of interesting results. First, we found that the water resource is better used under the MPMA in comparison to an unregulated monopoly. Second, a monopoly generator with flexible hydro can make a positive profit

under the MPMA in the case of a positive demand shock, while it makes zero profit when it has no water resource. Third, shifting the hydro resource between periods may be better than importing electricity in mitigating the price impact of an outage. Fourth, a marginal cost pricing rule imposed on a monopolist with both thermal and hydro generation will not lead to a socially optimal outcome.

## References

1. Aigner, D.J. and J.G. Hirschberg (1985) "Commercial/industrial Customer Response to Time-of-use Electricity Prices: Some Experimental Results", *Rand Journal of Economics* 16(3):341-355
2. Ambec, S. and J.A. Doucet (2003) "Decentralizing Hydro Power Production", *Canadian Journal of Economics* 36(3):587-603
3. Bailey, E. E. and L. J. White (1974) "Reversals in Peak and Off-peak Prices", *Bell Journal of Economics* 5:75-92.
4. Bushnell, J. (2000) "Water and Power: Hydroelectric Resources in the Era of Deregulation in the Western US", University of California Energy Institute, Discussion paper PWP-056.
5. Chute, Russell G. (2001) "The Impact of the Market Power Mitigation Agreement on Power Prices in Ontario", Insight Information Co. mimeo.
6. Goldman, C., N. Hopper, O. Sezgen, M. Moezzi, R. Bharvirkar, B. Neenan, R. Boisvert, P. Cappers, and D. Pratt (2001) "Customer Response to day-ahead Wholesale Market Electricity Prices: Case Study of RTP Program Experience in New York", Ernest Orlando Lawrence Berkeley National Laboratory, Research Paper LBNL-54761
7. Gravelle, H. (1976) "The Peak Load Problem with Feasible Storage", *Economic Journal* 86:256-277.
8. Heberlein, T. and G. Warriner (1983) "The Influence of Price and Attitude on Shifting Residential Electricity Consumption from On- to Off- peak Period", *Journal of Economic Psychology* 4(1-2):107-130
9. Laffont, J-J. and J. Tirole (1986) "Using Cost Observation to Regulate Firms", *Journal of Political Economy* 94: 614-641.
10. Laffont, J-J. and J. Tirole (1994) *A Theory of Incentives in Procurement and Regulation*, The MIT Press, Cambridge, Massachusetts.
11. Lewis, T.R. and D. Sappington (1988) "Regulating a Monopolist with Unknown Demand", *American Economic Review* 78(5): 986-998.
12. Littlechild, S. (2001) "Electricity: Regulatory Developments Around the World", The Beesley Lectures on Regulation Series XI.
13. MSP (Market Surveillance Panel), Market Monitoring Report, June 2003, December 2005, [www.oeb.gov.on.ca](http://www.oeb.gov.on.ca).

14. Panzar, J. C. (1976) "A Neoclassical Approach to Peak Load Pricing", *Bell Journal of Economics* 7:521-530.
15. Spulber, D. F. (1989) *Regulation and Markets*, The MIT Press, Cambridge, Massachusetts.
16. Steiner, P.O. (1957) "Peak-load and Efficient Pricing", *Quarterly Journal of Economics* 71:585-610.

# Chapter 4 Reliability and the MPMA

## 4.1 Introduction

Electric power systems are extremely complicated. By their nature they are exposed to many stochastic disturbances, such as generation failure (i.e. forced outages), transmission outages, and rapid consumption fluctuations (for example, due to a rapid change in temperature). Because electricity demand and supply must be balanced instantaneously, any unexpected fluctuations must be quickly responded to by the system operator, who could dispatch additional supply, shed load, purchase emergency energy from external jurisdictions, or take other out-of-market control actions (e.g. lower the operating reserve requirement or system voltage<sup>29</sup>). Because the market rules in Ontario limit the system operator's ability to take out-of-market control actions and because loads are unable to cut consumption (due to poor pricing signals/ incentives) or are technically unable to do so, a convenient approach to dealing with the supply / demand imbalance is to maintain spare generating capacity online so it can respond to the system operator's instructions instantaneously. If generators respond slowly, inappropriately, or insufficiently, system reliability will be threatened, or at a minimum, the system operator would fail in its reliability obligations to neighbouring jurisdictions by cutting exports. This could result in unstable voltage or frequency (which damages electric appliances), a blackout in certain regions, or in extreme cases, a collapse of the whole system (such as the one in Ontario and northeast USA on August 14, 2003).

---

<sup>29</sup> The system operator can reduce the system voltage by 3% or 5% under certain conditions to deal with system contingencies. Reducing the system voltage will lower power consumption although consumers may not notice it.

The most effective approach of maintaining reliability is to have spare generating capacity on line or in a ready-mode; this is called operating reserve. This operating reserve is either synchronized with the system with the potential to produce more electricity, or can be synchronized and produce power within 10 to 30 minutes.

Whenever there is an unanticipated disturbance on either the supply or demand side, those reserves can be quickly activated to generate power.

The definitions, functions and components of operating reserve are well documented in the Market Rules of the Independent Electricity System Operator<sup>30</sup>. Both the North American Electricity Reliability Council (NERC) and the Western Systems Coordinating Council (WSCC) define Operating Reserve in a similar way. They consider OR to be the capability in excess of firm system demand that is required to provide for regulation, load forecasting error, equipment forced outages, and local area protection.

The Federal Energy Regulatory Commission (FERC), however, seems to favour a narrower definition. In its Order No.888, OR is defined as the “extra generation available to serve load in case there is an unplanned event such as loss of generation”, which excludes the responsibility of OR for forecast errors. Although some researchers (Hirst and Kirby, 1997b, 1998) argue that the FERC has the correct definition because forecast error is purely commercial and thus the parties making the errors are responsible for them, it is practically impossible to penalize those who made forecast errors or reward those who made accurate forecasts. After all, in all deregulated markets, the central job

---

<sup>30</sup> For details, see: [www.ieso.ca](http://www.ieso.ca)

of forecasting is ultimately done by the system operators, who generally do not have the type of profit incentive to induce them to minimize forecast error (a system operator is a non-for-profit organization in all existing electricity markets).

There are two types of operating reserves: spinning and non-spinning reserves. Spinning reserve is online, synchronized, and ready to be activated on request to serve any demand or supply disturbance. In general, these resources can produce power within 10 minutes on request, and thus are called 10 minute spinning reserve. The non-spinning reserve usually falls into two categories: 10 and 30 minute non-spinning reserve. The 10 minute non-spinning reserve is comprised of generating capacity that is not connected to the system but is capable of synchronizing and generating power within 10 minutes, or those interruptible consumers (called dispatchable loads) that can be shed within 10 minutes. The 30 minute reserve is similar to the 10 minute non-spinning reserve, but over a longer response time horizon.

The minimum requirement for OR can be based on either deterministic or probabilistic calculations. The deterministic method has historically been applied in almost all interconnected systems because of its simplicity and practicability. This method requires that there be sufficient spinning OR to deal with the single largest contingency, or one and half times the largest single contingency event. These requirements are independent of the characteristics of generating units in the power system in question. For example, the N-1 standard (loss of a single transmission or generation element) requires the 10 minute spinning reserve to be the amount of the largest single on-line unit regardless of how well the unit performs (Ontario currently applies this standard).

The probabilistic approach is believed to be more consistent with the stochastic nature of supply disturbances (Joskow and Tirole, 2004, Hogan, 2005) and reliable because an electricity system is stochastic in nature. As far as we know, the Western Systems Coordinating Council (WSCC) is the only reliability council that takes into account the performance of different types of generation. Billinton and Allan (1996) further argue that there is no need to constrain artificially the inherent probabilistic or stochastic nature of a power system into a deterministic domain. They argue that the main reasons that lead to deterministic approach, including lack of data, limitation of computation resources, lack of realistic reliability techniques, aversion to the use of probabilistic techniques and a misunderstanding of the significance and meaning of probabilistic criteria and risk indices, are not valid today and thus a probabilistic approach should be doable.

This chapter shows that the probabilistic approach can actually be applied to construct a demand curve for OR which incorporates the stochastic nature of generating units and demand levels for energy, and that there are many benefits from doing so. Then, based on a generalized demand function for OR, a model is built to derive the optimal strategies to price both energy and OR. Although the model is based on the outage of generation units, the concept can be extended to load disturbances: an unexpected increase in demand is equivalent to a forced outage of generation units. In this chapter, we show that when supply uncertainty enters the supply curve, price should exceed the long run marginal cost in situations of both monopoly and competition (Carlton (1977) has an equivalent result in a stochastic demand setting).

This chapter is structured as follows. The next section describes the two approaches to determining OR requirements and their respective strength and weakness. The third section presents a literature review and comments. The fourth section illustrates a numerical example of the construction of a demand curve for OR. The fifth section builds a generalized model that connects the energy and OR markets. The sixth to tenth sections derive several results and discuss their policy implications. The last section presents some concluding remarks.

## **4.2 Approaches to Calculating Operating Reserve Requirements**

The deterministic approach has been used for more than 100 years and widely applied in almost all electricity control systems or markets. This approach was implemented through all stages of planning when the electricity generation was subject to regulation. At the stage of planning generating capacities, installed capacity is required to be equal to the expected maximum demand plus a fixed percentage. At the stage of planning operating capacities, spinning capacity must equal the expected load plus a reserve equal to the capacity of one or two largest generating units. At the stage of planning network capacities, a minimum number of circuits are constructed to serve a load group. With the deregulation of an electric power system, however, most if not all these planning functions disappear and the supply of electricity is mainly driven by market forces, which induce both private and public firms to invest or divest through market signals. Nevertheless, the deterministic method remains in the OR market, where the system operator typically determines the OR requirements to be equal to the capacity of one or

one and one-half of the largest on-line generating unit (e.g. in Ontario), or to a percentage of maximum expected load (e.g. in California).

The deterministic approach has been criticized for a long time because it fails to take into account other supply/demand parameters, e.g. unit-specific forced outage probabilities, the mix of online generating units, and heterogeneity of consumers. Specifically, the largest on-line unit may be very reliable such as a nuclear unit or a baseload hydro unit, and a requirement of spinning OR equalling the capacity of that unit may be too much. Individual consumers may put different reservation values on consumption interruption, and thus a single inelastic demand of OR may be inefficient. As a result, the OR requirements calculated using this method could lead to either excessive or insufficient levels of reliability (Joskow, 2005, Hogan, 2005).

In many markets where the deterministic approach is applied, the market operator faces a dilemma: typically when the OR is needed the most, its supply is not sufficient, while when it is not so needed, it is over-supplied. For instance, when demand is low and for the most part only those highly reliable units (baseload units) are online, the OR requirement is easy to meet but seems over-supplied. In contrast, when demand is high and thus more unreliable units are online, or during morning load pick-up when forced unavailability can be higher than normal for units just starting, the OR requirement is difficult to meet because of capacity or ramping limitations.

The probabilistic approach is based on the joint probability of unit failures. The required amount of on-line capacity is an increasing function of the forced outage rate of synchronized generation. If on-line generation consists of a large number of units with a high probability of tripping, the OR requirement needs to be higher so that a sudden loss of one or more generating units can be quickly offset by the stand-by reserves. Billinton and Allen (1996) have a detailed discussion of all aspects of electricity system reliability, which features abundant numerical examples showing that the probabilistic approach is possible and superior to the deterministic approach.

Of course, the probabilistic approach has its own shortcomings: it requires high quality data and powerful computational tools, both of which need a good deal of effort to either collect or design. Given that the computer technology has been significantly advanced and high quality data can be derived from the electricity industry which already has a history of more than 100 years, the probabilistic approach has become workable.

A simple numerical example may illustrate how the deterministic and probabilistic approaches can lead to different market outcomes. Assume the system exists for one period only and all generating units are identical with capacity 10MW. The forced outage rate (*FOR*) is 0.05 (i.e. the unit will have a process problem with a probability of 5 per cent). Also assume the forced outage of a unit is a 'sudden death', i.e. the whole capacity of the unit will trip completely. All other reliability failures (for example transmission line outages) are ruled out. The market demand is 100MW, and there are enough generators to provide this demand plus any amount of OR designated below.

Table 4.1 shows that the probability of demand being unsatisfied, i.e. the Loss of Load Probability (LOLP) (in other words, a blackout<sup>31</sup> in some area or all areas), falls rapidly as more generating capacity is kept online. The first column is the increment of spinning OR. The second column is the LOLP. The third column indicates the amount of OR required to meet the largest contingency event. The fourth column indicates the amount of OR required to cover 20 percent of demand. The fifth column is the NERC's reliability standard, i.e. the '1-day-in-10-years' standard for demand being greater than supply. This standard is equivalent to a blackout possibility of 0.03 per cent.

**Table 4.1: Operating Reserve and Reliability Levels**

Operating reserve (MW)	LOLP	Largest contingency event (10MW)	20 per cent of demand (20MW)	NERC's standard: 0.03%
0	0.401263			
10	0.101895	*		
20	0.019568		*	
30	0.003103			
40	0.000428			*
50	0.000053			
60	0.000006			
70	0.000000			
80	0.000000			

The formula for calculating the LOLP is simply a sum of the joint probabilities of being at least one unit shy of meeting energy demand, and can be expressed as follows:

$$LOLP = \sum_{i=0}^9 \binom{n}{i} FOR^{(n-i)} (1 - FOR)^{(i)} \quad (4.1)$$

where  $n = (100 + OR) / 10$ , the number of total online generating units.

<sup>31</sup> This is a simplified assumption because one or two units with outage may simply lead to 'brownout' (i.e. a drop in voltage or frequency), not necessarily to a blackout or rotating blackout.

One can see that reliability is lower than the NERC standard in the hypothetical case in which the OR requirement is the maximum lost MW of the largest contingency event. In particular, the possibility of demand being unsatisfied is 0.101895 if the N-1 standard is established, while the NERC reliability criterion requires a probability of 0.0003, which is equivalent to a reserve requirement of 4 units. In this case, the two standards set up by the NERC are inconsistent, as Joskow (2005) has observed. On the other hand, if the unit size is greater or the forced outage rate is smaller, the deterministic approach could result in realized reliability in excess of the probability standard – ‘1-day-in-10-years’.

#### **4.3 A Brief Review of Literature on Pricing Operating Reserve**

There is a great deal of literature dealing with the optimal level of or cost allocation of OR for an integrated utility before electricity deregulation. Under the old system of ‘cost-of-service’ or ‘rate-of-return’ regulation, the optimal pricing of OR is essentially the cost of keeping those reserves on-line to supply OR. A typical model for this system assumes a vertically integrated monopoly maximizing social welfare on behalf of the central planner. Then, based on the estimated cost of lost load and LOLP, an optimal provision and pricing of OR is determined (Telson, 1975, Crew and Kleindorfer, 1976, 1978, Siddiqi and Baughman, 1995). Because the costs of providing OR are embedded in the ‘bulk’ power price, consumers cannot identify how much they have paid for this specific service, and thus it is hard to say that this ‘optimal’ pricing is really optimal when consumers’ preferences are excluded. Furthermore, a deterministic OR requirement may lead to artificially higher costs to consumers. Cramton and Stoft (2006)

observe that the capacity reserve margin criterion used in the northeast USA reflects a value of lost load of US\$267,000/MWh, suggesting a much too high reserve margin. Joskow (2005) argues that implicit value of lost load associated with the NPCC's standard of one-day-in-ten-years is very high (\$300,000/MWh).

In recent years, the probabilistic approach has attracted a lot of attention from both engineers and economists, and has also been partially applied in California by WSCC. Many probabilistic models were built before the dramatic restructuring of electric power systems began in North America. Most of these models (e.g. Crew and Kleindorfer, 1976, 1978) were to deal with a vertically integrated monopoly, which was assumed to maximize social welfare on behalf of the government. These models appear inappropriate in a decentralized world.

There are few publications regarding OR markets, partially because the OR market is small in size, low in price and only an ancillary or derivative market. In recent years, some researchers started to turn their attention to the OR market, realizing its importance in joint optimization and in maintaining a reliable energy market. Kirby and Hirst (2003) and Hirst and Kirby (1997b) argue that the current approach of OR cost allocation is unfair and inefficient because those responsible for the forced outages or forecast errors don't have to pay for their inefficiency or mistakes. They argue that because the reserve cost is currently allocated to all consumers in proportion to their energy consumption rather than to generators, those poorly-performed generators (i.e. those with a high incidence of forced outages) have no incentive to reduce their outage rate. Then they

suggest that reserve costs should be allocated among those responsible according to their performance, although the final bill would ultimately be paid by consumers.

The Hirst and Kirby model fails to answer an important question: why some units have a higher outage rate than others. If this question is not properly answered, their proposal could actually lead to more problems than it solves. Under a fully integrated monopoly, a unit manager may have little incentive to improve his unit's performance because all costs will be covered under cost-plus or rate-of-return regulation. In a competitive market, however, a unit manager has strong incentives to reduce his outage rate because he will lose revenue if his unit trips off-line. Therefore any extra penalty on poor performance, such as the suggested reserve charge, has very limited impact on the manager's effort to improve the unit performance. Second, a forced outage is most likely to occur during the ramp-up period, which implies that a peaking unit, which has to start up frequently, will be more likely to have a high forced outage rate. High reserve charges on these peaking units could induce them to reduce start-up frequency and thus reduce energy supply when needed, or extend operating time by competing against base load suppliers, both of which are obviously inefficient. Alternatively, a peaking unit may increase its offer price so as to include a risk premium for a potential outage. As a result, the market clearing prices are likely to be higher.

Furthermore, there is practical difficulty implementing Hirst and Kirby's model. In their model the energy producers are assumed to buy insurance against the failure of generating machines from insurers who own backup capacity. However, there are in fact

opportunity costs to those insurers themselves, because they could physically bid their capacities into the energy market to earn revenue from it, rather than supplying OR. As a result, the insurers would charge a premium that is equal to their opportunity cost, which is positively related to the energy price. Given that the energy price is very volatile, the premium thus may change frequently, depending on the time of supplying OR, which eventually leads to large transaction costs. A large firm may find it beneficial to supply OR itself rather than purchase OR from a third party. Thus only small generators will have to go into the insurance market, if it exists.

Chan, Li, and Wu (2002) propose instead an insurance-like mechanism for consumers and argue that consumers' involvement in deciding the optimal level of OR would improve efficiency. This is true in a pure *private* good market without any externalities and other market failures; the 'invisible hand' can generally lead to the most efficient outcome when the market is competitive. In the electricity system, however, the OR requirement (or broadly speaking reliability) is actually a *public* good; it benefits all consumers regardless of who supplies it. Therefore the classic 'free-rider' problem appears: every consumer will try to avoid paying for OR supply because he/she can benefit from others paying. As a result, consumers may find it beneficial to buy insurance against loss of electricity supply rather than purchasing OR. Therefore, it may be better for a central planner, who works on behalf of all consumers, to bargain with generators to purchase sufficient OR.

If generators were responsible for the insurance based on their historical performance, the situation goes back to the Hirst and Kirby model, which is subject to other problems, as has been discussed before.

There do not appear to be any published articles modeling the interdependence of the energy and OR markets in a game-theoretic context (Joskow and Tirole (2004) is the first working paper on this issue). Some articles, e.g. Siddiqi and Baughman (1995), Chan, Li, and Wu (2002), model optimal energy and OR prices by assuming an exogenous, rather than endogenous OR demand, which completely ignores the dependence of OR on energy.

#### 4.4 A Demand Function for Operating Reserve

As shown in the previous numerical example, the LOLP decreases as online operating reserve capacity increases. This relationship could provide the basis for a demand curve for OR, if the LOLP can be linked to the willingness-to-pay for avoiding an involuntary curtailment of consumption<sup>32</sup>.

Fortunately, this is possible. The *Value of Lost Load* (VOLL) is the value that a consumer puts on interrupted consumption of electricity<sup>33</sup>, which is usually much higher

---

<sup>32</sup> In the Ontario electricity market, the demand curve for OR is step-wise. When the OR price reaches \$30, the demand for OR can be reduced up to 400MW; when the price reaches \$75, the OR requirement can be further reduced by 200MW; and when the price reaches \$100, the requirement can be lowered by another 200MW. However, this demand curve is based on historical OR prices and has nothing to do with the characteristics of on-line generation.

<sup>33</sup> VOLL is the marginal value of a lost Mw of consumption, representing a consumer's willingness-to-pay for avoiding an involuntary consumption reduction. An extremely high VOLL doesn't mean that consumer's demand for electricity is totally inelastic.

than the price of electricity that consumers are paying. To ensure his/her electricity consumption not to be interrupted, a typical consumer is thus willing to pay a premium for improving the system reliability, i.e. reducing the LOLP till the premium equals the marginal benefit.

In fact, a VOLL is difficult to estimate precisely because it depends on the types of consumer, the time of loss, and the non-linear dependence of lost value on the duration of the loss (Stoft, 2002). A wide range of estimates of VOLL (from US\$2,500 to \$20,000) is justified by some researchers (Willis and Garrod, 1997, Kariuki and Allan, 1996, Wacker and Billinton, 1989). In general, residential consumers have a relatively low VOLL, industrial consumers a high VOLL, and commercial consumers in between. Beenstock, Goldin and Haitovsky (1998) conduct a survey on household consumption of electricity, and find that the average outage cost is slightly higher than CDN\$7,000/MWh in 1990 dollars, depending on the outage time and duration, season, time-of-day, and day-of-week. Regardless of the apparent difficulties of estimating the precise VOLL, the upper bound of these estimates appears to be much less than the implied VOLL by the NPCC's reliability standard.

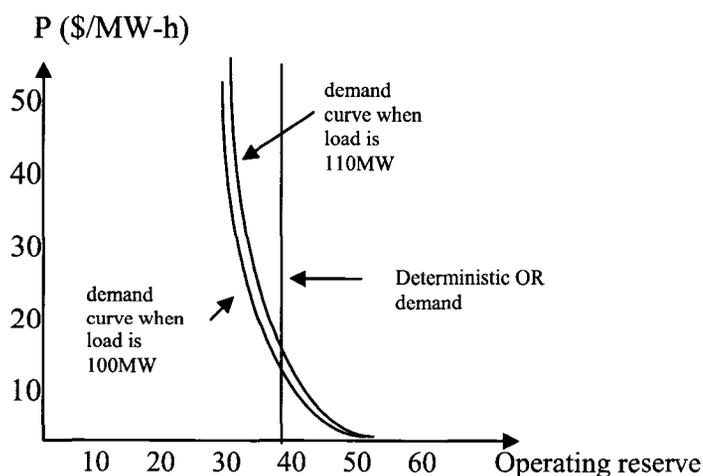
The VOLL concept has been applied in a competitive electricity market. In the early UK power market, the mechanism for the capacity payments was based on the VOLL. In April 2000, the VOLL was only £2816/MWh (Sweeting, 2004). A generation facility thus can be paid for its installed capacity based on the possibility of a blackout.

Table 4.2 below illustrates results based on the previous numerical example and an assumed average VOLL of \$20,000/MWh. Columns 1 and 3 (and 5) construct the demand curve for OR when energy demand is 100MW (and 110MW, respectively). Figure 4.1 reproduces these results. In equilibrium, the price of OR must be equal to VOLL times the LOLP in the sense of Joskow and Tirole (2004). Thus when the price for OR is lower, consumers are willing to purchase more OR.

**Table 4.2: Demand for Operating Reserves**

Operating reserve (MW)	LOLP ( $a$ ) (D1=100MW)	VOLL* $a$ (\$/MW-h)	LOLP ( $a$ ) (D2=110MW)	VOLL* $a$ (\$/MW-h)
0	0.401263	8025.26	0.431200	8624.00
10	0.101895	2037.90	0.118360	2367.2
20	0.019568	391.36	0.024508	490.16
30	0.003103	62.06	0.004174	83.48
40	0.000428	8.56	0.000615	12.30
50	0.000053	1.06	0.000081	1.62
60	0.000006	0.12	0.000010	0.20
70	0.000000	0.00	0.000001	0.02
80	0.000000	0.00	0.000000	0.00

**Figure 4.1: Operating Reserve Demand Curves**



The demand for OR is positively related to the outage rate. As the unit forced outage rate increases, the power system becomes less reliable, and consumers are willing to pay more for operating reserve.

Therefore, we can now establish a general demand function for OR, which depends on the forced outage rate and the energy demand level. In particular, we establish the following inverse OR demand function:

$$P^{OR} = f(q^{OR}, q^E, p^F) \quad (4.2)$$

where  $P^{OR}$  denotes the price for OR,  $q^{OR}$  the OR demanded,  $q^E$  the load level, and  $p^F$  the forced outage rate of a typical unit. Note that the demand for OR must be lower than the demand for energy because the forced outage rate is generally low<sup>34</sup>.

The OR demand function has the following characteristics:

$$\begin{aligned} \partial f / \partial q^{OR} < 0; \quad \partial^2 f / \partial (q^{OR})^2 > 0; \quad \partial f / \partial q^E > 0; \\ \partial f / \partial p^F > 0; \quad \frac{\partial(\partial f / \partial q^{OR})}{\partial q^E} < 0; \quad \frac{\partial(\partial f / \partial q^{OR})}{\partial p^F} < 0 \end{aligned}$$

The first two inequalities imply a convex demand curve. The third and fourth imply that the demand for OR is positively related to the load level and the forced outage rate. The fifth and the sixth imply that the OR demand curve becomes steeper as the demand for electricity increases, and/or as the forced outage rate increases. The reason for this is that a higher load requires more units on-line, for a given forced outage rate, leading to a higher probability of a blackout. Therefore, given any OR level, consumers are willing to

---

<sup>34</sup> There is no information about this in the public domain. According to industry insiders, the forced outage rate for a hydroelectric unit is less than 10 percent. A fossil unit is estimated to be about 20 percent.

pay more to secure the supply of electricity. Consumers' willingness to pay for additional MW of OR drops faster as more units come online to supply OR.

Holding  $P^{OR}$  constant and taking the total differential of equation (4.2) yields:

$$0 = \frac{\partial f}{\partial q^{OR}} * \Delta q^{OR} + \frac{\partial f}{\partial q^E} * \Delta q^E + \frac{\partial f}{\partial p^F} * \Delta p^F \quad (4.3)$$

Given the characteristics of the inverse demand function, the direct demand function,  $q^{OR}$  has following features:

$$\frac{\partial q^{OR}}{\partial q^E} = -\frac{\partial f / \partial q^E}{\partial f / \partial q^{OR}} > 0 \quad \text{and} \quad \frac{\partial q^{OR}}{\partial p^F} = -\frac{\partial f / \partial p^F}{\partial f / \partial q^{OR}} > 0$$

That is, the demand for OR increases as the demand for energy or the unit forced outage rate increases.

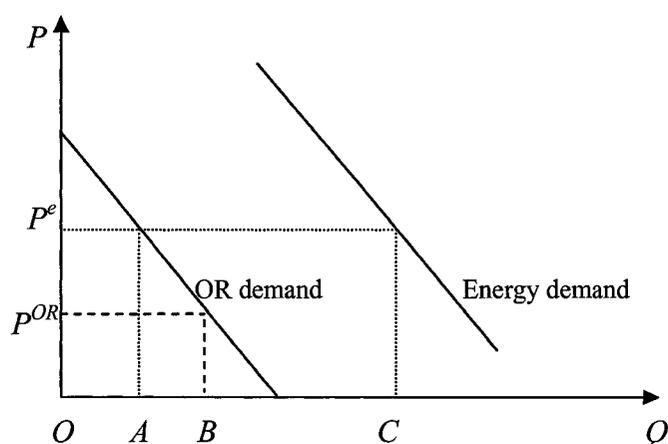
These results reveal significant differences between the probabilistic and deterministic approaches. As shown before, the deterministic approach doesn't take into account the demand for energy (or equivalently, the number of on-line generation units), or the probability of a forced outage. That is, the deterministic approach assumes

$$\frac{\partial q^{OR}}{\partial q^E} = 0 \quad \text{and} \quad \frac{\partial q^{OR}}{\partial p^F} = 0.$$

Because the inverse OR demand curve is relatively steep (i.e. the probability of sufficient unit failures drops dramatically as more OR supplying units are dispatched on line), it is reasonable to assume that the impact of energy demand on the OR price is greater than on the OR quantity, i.e.  $\frac{\partial P^{OR}}{\partial q^E} > \frac{\partial q^{OR}}{\partial q^E}$ .

Chan, Li, and Wu (2002) construct a benefit function of OR rather than an explicit OR demand function, while Borenstein and Holland (2003) assume the demand for OR to be a fixed proportion of energy demand. The first approach generally ignores the interaction between the demand for energy and the demand for OR. The second approach mimics the practice of requiring more OR during the peak hours and less OR during the off-peak hours, but the function of OR demand constructed in this way will not reach the equilibrium as intended, except in the case of vertical energy demand. To see this, refer to Figure 4.2 below, where the demand for OR is constructed as a fraction of the energy demand at any given energy price. Assume that the equilibrium energy price is  $P^e$  and energy demand is  $OC$ . Then the required OR is  $OA$ , i.e. a fraction of  $OC$ . Because the cost of supplying OR is lower than the cost of supplying energy, the actual OR supply will be  $OB$  in a competitive market, which is greater than  $OA$ . That is, the equilibrium OR supply will be larger than the target when the demand curve is sloping down.

**Figure 4.2: Demand for OR Is Proportional to the Demand for Energy**



Deriving the correct OR demand function is thus important for finding the socially optimal OR requirement and the optimal level of investment in capacity.

#### 4.5 The Model

Assume that the demand for energy depends on the energy price only, denoted as  $P^E = g(q^E)$ , which has characteristics  $dg(q^E)/dq^E < 0$  and  $d^2g(q^E)/dq^{E2} > 0$  ( $d^2g(q^E)/dq^{E2} = 0$  in a linear demand model).

A generating unit is assumed not to provide energy and OR at the same time<sup>35</sup>, and operates at full capacity upon dispatch (equivalent to the assumption that each single plant can produce only one unit of either energy or OR as in Borenstein and Holland (2003)). All units have identical cost functions  $C_i(q_i^E) = F + bq_i^E$  for energy production and  $C_i(q_i^{OR}) = F$  for OR supply, where both  $q_i^E$  and  $q_i^{OR}$  are equal to the generating unit's maximum capacity  $\bar{q}$ . In other words, each unit has a fixed cost  $F$ , marginal cost  $b$  for producing a unit of energy, and zero marginal cost for OR. In the short-run,  $F$  doesn't affect a generator's production decision.

The monopolist is risk neutral and thus maximizes total profits from both the energy and the OR markets. Unlike a firm in a competitive market, who has to directly incorporate the possibility of unit failure into her supply strategy, the monopolist doesn't have to do so. The reason is that the demand curve for operating reserve has already factored in all

---

<sup>35</sup> In reality, a unit can provide both energy and spinning OR at the same time, subject to its ramp rate and capacity level.

possible scenarios of system failure. Additional consideration of outage by the monopolist simply leads to double counting. For example, if the monopolist takes into account the possible failure of units in the energy market by supplying more stand-by reserve, the possibility of system collapse will decrease and thus the demand for OR will drop. This drop in demand for OR will lower the monopolist's profit in the OR market. As a result, the monopolist will not directly consider the outage issue in its supply functions.

The different treatment of the unit outage by a monopolist and a competitive firm reflects two fundamentally different views: the monopolist views the energy and OR market as substitutes and complements at the same time, while a competitive firm sees them as substitutes only.

Because each unit produces at its full capacity, choosing MW in either energy or OR is equivalent to choosing the number of units to be online. Thus the monopolist's problem becomes:

$$\begin{aligned} \underset{M,N}{Max}: & [g(q^E)q^E - NF - Nb\bar{q}] + [f(q^{OR}, q^E, p^F)q^{OR} - MF] & (4.4) \\ s.t.: & (1) \quad q^E = N\bar{q} \\ & (2) \quad q^{OR} = M\bar{q} \end{aligned}$$

The first term in the objective function is the profit in the energy market and the second term the profit in the OR market.  $N$  is the total number of units supplying energy, and  $M$  the total number of units supplying OR.

Then the FOC for the above problem are:

$$g(N\bar{q}) + N\bar{q} \frac{dg(N\bar{q})}{dN} - \frac{F}{\bar{q}} - b + M\bar{q} \frac{\partial f(M\bar{q}, N\bar{q}, p^F)}{\partial N} = 0 \quad (4.5a)$$

$$f(M\bar{q}, N\bar{q}, p^F) + M\bar{q} \frac{\partial f(M\bar{q}, N\bar{q}, p^F)}{\partial M} - \frac{F}{\bar{q}} = 0 \quad (4.5b)$$

The first FOC says that the marginal revenue of an additional plant to supply energy should be equal to the average cost of producing energy minus *cross* marginal revenue from energy production in the OR market. In contrast, in an energy only setting, the FOC is  $g(N\bar{q}) + N\bar{q} \frac{dg(N\bar{q})}{dN} - \frac{F}{\bar{q}} - b = 0$  assuming there are no outages. Thus, (4.5a) has an additional term  $M\bar{q} \frac{\partial f(M\bar{q}, N\bar{q}, p^F)}{\partial N}$ , which is positive, implying that the monopolist is now producing more energy. The intuition here is that by producing more energy, the monopolist puts more generating units online for energy, and thus increases the possibility of a system outage and thus the demand for OR. As a result, the monopolist makes a higher profit in the OR market, which is more than its lost profit in the energy market.

In other words, the energy and OR supplies are one-direction complements in this situation: the more energy is produced, the more OR is needed. Because the monopolist can make profits in both complementary product markets, joint optimization will lead to higher quantities produced in two markets and thus lower prices than would be if the two markets were isolated. As we will see later, however, energy and OR could become substitutes when the monopolist has a capacity limit: if capacity is used to produce energy, it cannot be used to supply OR.

Given a forced outage rate and a unit maximum capacity, the FOC's can be solved for the two unknown variables  $M$  and  $N$ , and thus the equilibrium outputs, denoted  $q^{E*}(\frac{F}{q}, p^F)$  and  $q^{OR*}(\frac{F}{q}, p^F)$ , both of which are positively related to  $p^F$  and negatively to average fixed cost  $\frac{F}{q}$ .

If  $p^F$  is endogenous, that is, if the monopolist can claim  $p^F$  based on its own strategy, according to problem set (4.4), the monopolist will choose a highest  $p^F$  because the profit function is monotonically increasing in  $p^F$ . Therefore even the monopolist has incentives to falsely claim an outage. In this case, a truth telling mechanism should focus on the OR market although the joint optimization makes profit in the two markets interdependent. The reason for this is that the forced outage rate affects only the demand for OR, not the demand for energy. Given the asymmetric information on the true forced outage rate, a truth telling mechanism requires some regulation in the OR market.

The regulation would be very challenging when the monopolist has the ability to shift the demand curve rather than the demand level along a known demand curve. Most of the literature in information economics deals with asymmetric information on demand which is exogenous to the firm (adverse selection), or on cost which is endogenous to the managerial effort (moral hazard), or both. The forced outage issue can be seen as a variant of moral hazard, and the effort of maintaining a good performance of generating machines is unobservable. In this case, the contract given to the firm is necessarily a fixed price contract, and the regulator has no control over the power of any incentive scheme (Laffont and Tirole, 1994, Baron and Myerson, 1982).

Thus, to induce best performance or truth telling, the forced outage rate used to calculate the OR demand function must be independent of the claimed outage rate. This can be done through yardstick regulation, i.e. use observed forced outage rates on similar facilities in other competitive markets. Another approach is to contract out or auction off part or most of the OR supply to independent generators, which would significantly reduce the monopolist's incentive to falsely claim a higher outage rate.

Given that  $q^E = N\bar{q}$  and  $q^{OR} = M\bar{q}$ , FOC (4.5a) and (4.5b) can be rewritten in a more familiar form:

$$g(q^E) + q^E \frac{dg(q^E)}{dN} - \frac{F}{\bar{q}} - b + q^{OR} \frac{\partial f(q^E, q^{OR}, p^F)}{\partial N} = 0 \quad (4.5a')$$

$$f(q^{OR}, q^E, p^F) + q^{OR} \frac{\partial f(q^E, q^{OR}, p^F)}{\partial M} - \frac{F}{\bar{q}} = 0 \quad (4.5b')$$

Subtracting (4.5b') from (4.5a') and rearranging the equation yields the difference between the energy price and the OR price:

$$g(q^E) - f(q^{OR}, q^E, p^F) = b - q^E \frac{dg(q^E)}{dN} + q^{OR} \left( \frac{\partial f}{\partial M} - \frac{\partial f}{\partial N} \right) \quad (6)$$

On the right hand side, the first two terms are positive, while the last term is negative (because the own effect of OR supply (the impact of producing one more unit of OR on the OR price) is greater than the cross effect (the impact of producing one more unit of energy on the OR price)). If the OR demand curve is much steeper than the energy demand curve (e.g. as a result of a high forced outage rate), the right hand side will likely be negative. Thus the energy price is not necessarily greater than the OR price even if the marginal cost of providing OR is smaller (a price reversal effect). The implication is that

because the demand for OR is less elastic, the most efficient pricing strategy for the monopoly is to bid a higher OR price. Intuitively, the OR price depends on the VOLL, which is usually very high, while the electricity price reflects the consumers' willingness to pay when there is no system failure.

However, the energy price is greater than the OR price if  $\frac{q^{OR}}{q^E} < \frac{dg(q^E)/dN}{\partial f/\partial M - \partial f/\partial N}$  (i.e.

$\frac{M}{N} < \frac{dg(N\bar{q})/dN}{\partial f/\partial M - \partial f/\partial N}$ ), which is more likely to be the case the steeper the energy demand curve,

and/or the lower the forced outage rate. Given that the demand for energy is not very elastic, the energy price is thus generally greater than OR price but the difference between the energy price and the OR price won't be exactly the difference in marginal cost, which is equal to  $b$ .

The equilibrium OR supply  $q^{OR*}(\frac{F}{q}, p^F)$  could be either greater or smaller than the traditional deterministic standard where the OR requirement is equal to the capacity of one or two largest generating units. The probabilistic standard mainly depends on the unit characteristics and the demand level, while the deterministic standard relies solely on the capacity of specific units. Consequently, they are generally not comparable.

If the OR requirement is established through a deterministic approach by the system operator, for example, the OR requirement is  $q_d^{OR*}$ , the monopolist's problem becomes:

$$\begin{aligned}
\underset{M, N, q^{OR}}{\text{Max}}: & [g(q^E)q^E - NF - Nb\bar{q}] + [q_d^{OR*} p^{OR} - MF] & (4.7) \\
\text{s.t.}: & (1) \quad q^E = N\bar{q} \\
& (2) \quad q_d^{OR*} = M\bar{q}
\end{aligned}$$

where  $P^{OR}$  is the OR price that is independent of the OR supply. The equilibrium for this problem is very simple: the monopolist will maximize its profit through maximizing the profit in each market as if they were not linked, because there is actually no interaction between the two markets given that the demand in the OR market is perfectly inelastic.

In equilibrium, energy production  $q_d^{E*}$  is determined by  $g(q_d^{E*}) + q_d^{E*} \frac{dg(q_d^{E*})}{dq^E} - \frac{F}{q} - b = 0$ .

The energy production is lower compared to where the OR requirement is established based on the outage probability. Meanwhile, because the OR requirement is pre-determined and completely inelastic, the price of OR will be either infinite (or the OR price cap, if any), or zero. That is, a slight change in demand or supply can lead to a large swing in the OR price.

Therefore the deterministic approach will lead to lower energy production, and consequently higher energy (and OR) prices. This implication could be extended to an oligopoly model. Because the OR demand is completely inelastic, firms may find that competing in the energy market while colluding in OR market is mutually beneficial. The OR price, however, could go down to zero if the oligopolists are playing a non-cooperative game.

If the NERC reliability standard is arbitrarily set to be  $a=0.03\%$ , for example, the equilibrium will hinge on who, the suppliers or the buyers, has to meet the requirement.

If the consumers are required to buy this level of reliability at whatever OR price prevails, the demand for OR (denoted by  $q_a^{OR*}$ ) will be vertical and can be derived from  $f(q_a^{OR*}, q^E, p^F) = a * VOLL$ . This is equivalent to the deterministic approach.

In contrast, if the firm is required to meet the standard (here the NERC serves as the regulator), although consumers still pay the bill through the market determined OR price, the monopolist has to satisfy  $f(q^{OR}, q^E, p^F) = a * VOLL$  at an OR price equal to  $a * VOLL$ , which is the price that consumers are willing to pay for a specified OR level.

Then  $q^{OR} = q_a^{OR}(q^E, p^F, a, VOLL)$ , and the monopolist's problem will be:

$$\begin{aligned} \text{Max}_N \quad & [g(q^E)q^E - NF - Nb\bar{q}] + [a * VOLL * q_a^{OR}(q^E, p^F, a, VOLL) - MF] \quad (4.8) \\ \text{s.t.} \quad & (1) \quad q^E = N\bar{q} \\ & (2) \quad M = q_a^{OR}(q^E, p^F, a, VOLL) / \bar{q} \end{aligned}$$

The FOC for this problem is simply:

$$g(q^E) + q^E \frac{dg(q^E)}{dq^E} - \frac{F}{q} - b + a * VOLL * \frac{\partial q_a^{OR}(q^E, p^F, a, VOLL)}{\partial N} = 0 \quad (4.9)$$

Compared to (4.5a'), energy production could be lower under the pre-determined reliability standard if  $a * VOLL * \frac{\partial q_a^{OR}(q^E, p^F, a, VOLL)}{\partial N} < q^{OR} \frac{\partial f(q^{OR}, q^E, p^F)}{\partial N}$ , i.e. the cross marginal revenue of energy production in the deterministic regime is smaller than the cross marginal revenue in the probabilistic regime. This is likely the case when a high reliability standard is required (i.e. a smaller  $a$ ), which leads to  $\frac{\partial f(q^{OR}, q^E, p^F)}{\partial N}$  being greater

than  $\frac{\partial q_a^{OR}(q^E, p^F, a, VOLL)}{\partial N}$ . The intuition here is that to meet the high reliability standard, the monopolist has to put more OR resources online, which will reduce both price and marginal profit in the OR market, and thus the monopolist may produce less energy so that less OR capacity is needed.

Similarly, energy production could be lower under a pre-determined reliability standard if the VOLL is smaller<sup>36</sup>. Because the reliability standard is arbitrarily set, consumers may put a low value on it when the firm has to supply it. Thus, the monopolist cannot gain much from the OR market, and has little incentive to increase the supply of OR. At one extreme, the VOLL is zero, i.e. consumers don't care about reliability, then the OR market disappears and the monopolist maximizes profit in energy market only.

In the United States, some ISO's impose an extremely low price cap in certain areas on the OR market (e.g. New York sets a price cap of \$2.52/MW-h for spinning reserve in certain zones). Such a low price cap will significantly reduce profits in the OR market, and in fact make them negligible. Thus the problem becomes one of maximizing profit from the energy market only. It is apparent that the resulting energy output would be lower than that from joint optimization without the OR price cap, as shown before.

In summary, we have the following result 1.

---

<sup>36</sup> Beenstock, Goldin and Haitovsky (1998) conducted a survey on household consumption of electricity, and found that the average outage cost is slightly higher than Cdn \$7,000/MWh in 1990 prices, depending on the outage time, season, time-of-day, and day-in-week. Unfortunately, they didn't present an analytical framework on how the demand and reliability will interact.

**Result 1** *Without a capacity constraint, the monopoly equilibrium will have the following characteristics:*

1. *The energy price is not necessarily greater than the OR price although the cost of providing OR is lower than the cost of providing energy;*
2. *The monopolist will produce more electricity so as to increase the demand for operating reserve when a probabilistic OR pricing approach is applied;*
3. *The monopolist has incentives to overstate its forced outage rate. Given asymmetric information, incentive regulation would be associated with a reference forced outage rate or a fixed forced outage rate;*
4. *The deterministic approach to calculating the OR requirement may lead to either higher or lower reliability than the optimum reliability level; and the deterministic approach will lead to higher energy and OR prices, and lower energy production than the probabilistic approach;*
5. *Energy output could be lower than the unregulated monopoly output, if the monopolist is required to meet a high reliability standard; and*
6. *A low OR price cap will reduce the monopolist's energy production and thus increase the energy price.*

#### **4.6 A Capacity Constrained Monopolist**

Electricity markets often face resource constraints and on many occasions the system operators have to curtail consumption so that the existing capacity can meet demand. An example is California in 2000 when rolling blackouts occurred quite frequently during the

summer. In such tight supply conditions, the monopolist has to optimally allocate its total capacity between the energy and OR markets. Facing a capacity constraint, the monopolist's problem is the following:

$$\begin{aligned} \underset{M,N}{Max} : & [g(q^E)q^E - NF - Nb\bar{q}] + [f(q^{OR}, q^E, p^F)q^{OR} - MF] & (4.10) \\ s.t. : & (1) \quad q^E = N\bar{q} \\ & (2) \quad q^{OR} = M\bar{q} \\ & (3) \quad M + N \leq L \end{aligned}$$

where  $L$  is the maximum number of available units, which is smaller than the optimal number of units  $L^*$ .

By assumption, the size of each unit is identical. Then the constraint (3) in problem (4.10) can be rewritten as  $q^E + q^{OR} < L\bar{q}$ . The FOCs for the constrained monopoly problem are as follows:

$$g(q^E) + q^E \frac{dg(q^E)}{dq^E} - \frac{F}{q} - b + q^{OR} \frac{\partial f(q^{OR}, q^E, p^F)}{\partial N} - \lambda = 0 \quad (4.10a)$$

$$f(q^{OR}, q^E, p^F) + q^{OR} \frac{\partial f(q^{OR}, q^E, p^F)}{\partial M} - \frac{F}{q} - \lambda = 0 \quad (4.10b)$$

$$q^E + q^{OR} = L\bar{q} \quad (4.10c)$$

where  $\lambda$  is the Lagrange multiplier on the capacity constraint (or the shadow price of capacity), which is positive. As one might expect, both the energy and OR prices are higher than the unconstrained monopoly prices. This is a normal result with a capacity constraint.

However, the difference between the energy price and the OR price will be exactly the same as if there were no capacity constraints. To see this, simply subtract (4.10a) from (4.10b) and rearrange it, the result will be the same as equation (4.6). This means that

when facing a capacity limit, the monopolist will produce at full capacity but increase both the energy and the OR price by the same amount.

If the demand for OR is deterministically established and there is no price cap in the OR market or the price cap is very high, the monopolist will find it profitable to maximize profit in the OR market first and put the remaining resources in the energy market. If the demand curve for energy is steep, the energy price will be very high, which is politically expensive. “The absence of an appropriate operating reserve demand curve is one of the difficulties in market designs that result in de facto price caps and missing money” (Hogan 2005, p.13). Consequently, the OR price will be either infinite or equal to the price cap, and the OR supply will be equal to the full amount of the deterministic quantity. The remaining capacity will go to the energy market. In this situation system reliability is too high while energy is undersupplied. In contrast, if the OR price cap is too low, the monopolist will first maximize profit in the energy market, and then put all remaining capacity into the OR market. Then the system reliability may be too low, and energy is oversupplied.

### ***Result 2***

1. *when the OR requirement is probabilistically determined, a capacity constrained monopolist will increase both the energy and OR price by the same amount relative to prices when there is no capacity constraint,; or*
2. *when the OR requirement is deterministically established, a monopolist will supply the full amount of OR requirement, and put all remaining capacity in the*

*energy market, if there is no OR price cap or the OR price cap is very high. In this case, the reliability is too high; or it will generate electricity that maximizes profit in the energy market, and put all remaining capacity into the OR market, if the OR price cap is low. Thus the reliability level is too low.*

#### **4.7 First Best and Second Best**

As usual, the social optimum is defined as the equilibrium that maximizes the sum of consumer and producer surplus. In a world with a single plant and without machine failure, maximal social welfare occurs where the marginal cost and demand functions intersect.

However, the possibility of machine malfunction complicates the optimization, especially when there are multiple production units in the marketplace. In a single market, equating the market price to *physical* marginal cost, which ignores the cost of machine failure, could lead to a market price which is too low, because machine failure will impose direct or indirect costs on the owners<sup>37</sup>. Therefore, the possibility of machine failure must be taken into account in determining the socially optimal price.

Outages also make market power monitoring more difficult. For example, a popular monitoring practice for measuring a participant's strategy is to compare the participant's offers with the unit heat rate regardless of the possibility of unit trip. For reasons stated

---

<sup>37</sup> The cost resulting from a unit outage could include the direct cost such as the cost of repair, staffing cost, and restart cost, and indirect cost such as contract commitment and opportunity cost (for example if for environmental reasons a hydro unit has to flow water whether it can generate power or not).

above, this approach might lead the monitoring authority to over-react to a change in offer prices, increasing the possibility of false positives.

In the short run when the capacity is sufficient, the first best solution is marginal cost pricing. Thus, the short-run optimal production/consumption  $q_{Opt}^E = g^{-1}(b)$ , given the energy demand function  $P^E = g(q^E)$  and marginal cost  $b$  of producing energy.

Substituting the energy supply into the OR demand function and setting it equal to zero yields optimal OR supply  $q_{Opt}^{OR} = f^{-1}(g^{-1}(b), p^F)$ . In equilibrium, the energy price is  $b$ , the energy consumption  $g^{-1}(b)$ , the OR price is zero, and the demand for OR  $f^{-1}(g^{-1}(b), p^F)$ .

However, the above short term equilibrium may not be socially optimal in the long run if the fixed cost of adding the last generating plant is greater than its benefit to society. In particular, the investment of last plant will incur a fixed cost  $F$ , which could be greater than its contribution to consumer surplus. In this case, either the energy price or the OR price should be higher than marginal cost.

The impact of energy on OR is in a single direction. If one unit is removed from the energy market, for example, the existing OR supply would be too much as a result of  $\frac{\partial q^{OR}}{\partial q^E} > 0$ . If one unit is removed from the OR market, however, there will be no impact on the energy market (no feedback effect). Thus in the long run equilibrium

model, we can solve the optimum by finding the optimal capacity in the energy market first, and then the optimal OR capacity based on the optimal energy capacity supply.

In the energy market, a unit should not be built if the gain (producer plus consumer surplus) is smaller than the fixed cost, that is, if

$$\int_{q_s^{E^*}}^{q_s^{E^*} + \bar{q}} (g(q) - b) dq < F \quad (4.12)$$

where  $q_s^{E^*}$  is the minimum value that can support the inequality (4.12). Because the energy demand function is well-behaved, i.e.,  $dg(q^E)/dq^E < 0$  and  $d^2g(q^E)/dq^{E^2} > 0$ , there must exist a minimum  $q$ , denoted as  $q_s^{E^*}$ , that satisfies (4.12). As a consequence, the corresponding energy price will be  $g(q_s^{E^*})$ , which must be greater than  $b$ . Given the unit capacity  $\bar{q}$ , the socially optimal number of plants supplying energy is  $q_s^{E^*} / \bar{q}$ . Apparently all units can recover their cost and even make a little profit. Thus the second best is also the first best.

Given that the optimal energy demand is known, the demand for OR is thus  $f(q^{OR}, q_s^{E^*}, p^F)$ . As result, a unit should not be built if

$$\int_{q_s^{OR^*}}^{q_s^{OR^*} + \bar{q}} f(q, q_s^{E^*}, p^F) dq < F \quad (4.13)$$

where  $q_s^{OR^*}$  is the minimum value that can support the inequality (4.13). Similar to the energy demand function, the OR demand function is well behaved as well, and thus there must be a minimum  $q$ , denoted as  $q_s^{OR^*}$ , to solve constraint (4.13). The socially optimal OR price will be  $f(q_s^{OR^*}, q_s^{E^*}, p^F)$ , which is equal to or greater than zero. Given the unit

capacity  $\bar{q}$ , the socially optimal number of plants supplying OR is  $q_s^{OR*} / \bar{q}$ . Therefore the optimal number of plants is  $(q_s^{E*} + q_s^{OR*}) / \bar{q}$ . Again all units can recover their cost.

Similar to the monopoly case, the social planner will not consider the outage issue in her supply function again because the demand for OR has already incorporated all possibilities of system failure. Thus, the model of social welfare optimization with multiple plants has no fundamental difference from the case of a single plant. The minor difference is that in the multi-plant case the social optimum needs to take into account the fixed cost of additional plants (integer problem), while in the single plant context the fixed cost is irrelevant to the equilibrium.

The above analysis can be summarized as follows.

**Result 3 (First Best and Second Best)** *The socially optimal price is not necessarily equal to marginal cost in the long run. In particular, no new investment should be added if*

*either  $\int_{q_s^{E*}}^{q_s^{E*} + \bar{q}} (g(q) - b) dq < F$  or  $\int_{q_s^{OR*}}^{q_s^{OR*} + \bar{q}} f(q, q_s^{E*}, p^F) dq < F$ , where  $q_s^{E*}$  is the minimum*

*value that can support the inequality and  $q_s^{OR*}$  is the minimal value that can support the*

*inequality. Then the optimal energy price and supply will be  $g(q_s^{E*})$  and  $q_s^{E*}$ ,*

*respectively. The optimal OR price and supply will be  $f(q_s^{OR*}, q_s^{E*}, p^F)$  and  $q_s^{OR*}$ ,*

*respectively. If  $q_s^{E*}$  happens to equal  $q_{Opt}^{E*}$ , the socially optimal energy price will be  $b$ .*

*Further, if  $q_s^{OR*}$  happens to equal  $q_{Opt}^{OR*}$ , the socially optimal OR price will be zero.*

#### 4.8 The Optimal MPMA

The MPMA, when applied to the energy market only, will have an impact on both the energy and OR markets because a binding MPMA will lead to higher energy production compared to an unconstrained monopoly, as shown in previous chapters. Higher energy production, in turn, will increase the demand for OR. In a single energy market with a single production unit and with outage possibility, as shown before, the optimal MPMA is to set the price cap to be the marginal cost and  $CRQ$  to be the demand level at the price cap. Here we will discuss the optimal MPMA that maximizes aggregate welfare in both energy and OR markets.

The previous section shows that the social optimum doesn't necessarily imply marginal cost pricing in a multi-plant industry with a plant-specific fixed cost. Intuitively, the social optimum should imply an optimal MPMA with the  $CRQ$  being equal to  $q_s^{E*}$  and a price cap  $g(q_s^{E*})$ .

In the presence of this "optimal" MPMA, the monopolist's problem now becomes:

$$\begin{aligned} \underset{M,N}{Max} : & [g(q^E)q^E - NF - Nb\bar{q}] + [f(q^{OR}, q^E, p^F)q^{OR} - MF] - rebate & (4.14) \\ s.t. : & (1) \quad q^E = N\bar{q} \\ & (2) \quad q^{OR} = M\bar{q} \\ & (3) \quad rebate = \begin{cases} 0, & \text{if } g(q^E) \leq P_{cap} \\ (g(q^E) - P_{cap}) * CRQ, & \text{if } g(q^E) > P_{cap} \end{cases} \end{aligned}$$

It is straightforward that the monopolist will not produce energy more than  $q_s^{E*}$ . On the other hand, the monopolist will not produce less than this quantity either because the

energy price would otherwise be greater than the price cap. The FOC then becomes

$$g(q^E) + (q^E - q_s^{E*}) \frac{dg}{dq} - \frac{F}{q} - b + q^{OR} \frac{\partial f}{\partial N} = 0.$$

Recall that (in Chapter 2) the market equilibrium price would be the price cap if the marginal revenue at  $CRQ$  is equal to or greater than the price cap. That is, the monopolist will bid at the price cap

$$\text{if } g(q^E) + (q^E - q_s^{E*}) \frac{dg}{dq} + q^{OR} \frac{\partial f}{\partial N} \geq \frac{F}{q} + b \text{ at } q^E = q_s^{E*}.$$

This is true because  $g(q_s^{E*})$  must be greater than or equal to  $\frac{F}{q} + b$  according to the social optimum,  $(q^E - q_s^{E*}) \frac{dg}{dq} = 0$ ,

and  $q^{OR} \frac{\partial f}{\partial N} \geq 0$ . As a result, under the optimal MPMA the monopolist will produce

energy at the optimal level, which implies an improvement in social welfare even with MPMA applied to the energy market only.

However, the OR market is not subject to the MPMA, implying that the firm will earn an unconstrained monopoly profit in the OR market. Given the energy demand level  $q_s^{E*}$ , the OR demand function  $P^{OR} = f(q^{OR}, q_s^{E*}, p^F)$  is established. Then the monopoly OR supply is implied by FOC  $f(q^{OR}, q_s^{E*}, p^F) + q^{OR} \frac{\partial f}{\partial q^{OR}} = 0$ . Of course, the monopoly OR supply must be lower than the social optimum. Therefore, in a for-profit monopoly the reliability of the electricity supply is lower than the social optimum. Yet this OR supply is still greater than the unconstrained monopoly supply because the MPMA has induced a higher energy production and, in turn, higher demand for OR.

The above analysis is based on the assumption that the monopolist has sufficient capacity to supply the energy and OR markets simultaneously. However, during some periods with extremely high demand, supply may be insufficient to serve both markets. In this

case, the system operator must either shed load, or purchase emergency energy from neighbouring markets, or reduce the OR requirement (when the OR requirement is deterministically established). Because all these approaches are out-of-market actions, we are not going to discuss their impact on market equilibrium. Instead, we will only assess the shortage condition in a market framework.

When capacity is limited, the monopolist's problem is as follows:

$$\begin{aligned} \underset{M,N}{Max} : & \{g(q^E)q^E - NF - Nb\bar{q}\} + \{f(q^{OR}, q^E, p^F)q^{OR} - MF\} - rebate & (4.15) \\ s.t.: & (1) \quad q^E = N\bar{q} \\ & (2) \quad q^{OR} = M\bar{q} \\ & (3) \quad M + N \leq L \\ & (4) \quad rebate = \begin{cases} 0, & \text{if } g(q^E) \leq g(q_s^{E*}) \\ (g(q^E) - b) * q_s^{E*}, & \text{if } g(q^E) > g(q_s^{E*}) \end{cases} \end{aligned}$$

The FOCs solving the constrained monopoly problem are:

$$g(q^E) + (q^E - q_s^{E*}) \frac{dg(q^E)}{dq^E} - \frac{F}{q} - b + q^{OR} \frac{\partial f(q^{OR}, q^E, p^F)}{\partial N} - \lambda = 0 \quad (4.15a)$$

$$f(q^{OR}, q^E, p^F) + q^{OR} \frac{\partial f(q^{OR}, q^E, p^F)}{\partial M} - \frac{F}{q} - \lambda = 0 \quad (4.15b)$$

$$q^E + q^{OR} = L\bar{q} \quad (4.15c)$$

where  $\lambda$  is the shadow price of capacity constraint.

Marginal profit is now equal to the shadow price of the capacity constraint, and the firm is producing less electricity than the  $CRQ$ . As a consequence, the firm has to pay a refund to consumers.

In equilibrium, the energy price will be greater than the price cap, the energy supply is less than the  $CRQ$ , and the price difference between the energy and OR prices is smaller

than it would have been without a capacity constraint. The implication is that once the price cap is binding, the firm has less incentive to bid a high price in the energy market because of the rebate mechanism, thus the price difference between two markets will be smaller.

In the Ontario market, the system operator (IESO) determined that the OR market was not competitive<sup>38</sup> when the market opened and required OPG to submit suitable offers to the OR market. The IESO, however, doesn't have the authority to regulate OPG's offer price; thus the OPG can bid any price as long as the offer is below the market cap (\$2,000/MWh). The market situation has been largely improved as more generators, dispatchable loads, importers and exporters enter the OR market. According to our earlier analysis, OPG has strong incentives to bid a high OR price because it is not subject to a rebate mechanism. If the OR market is more competitive than the energy market (due to the above reasons), a high OR price would induce more competition, and thus the dominant generator would have a lower market share in the OR market.

We have already seen the impact of the MPMA when applied to the energy market only. A logical question is then: what happens if the MPMA is also applied to the OR market? What is the optimal MPMA?

---

<sup>38</sup> Just like the energy market that was determined by the Market Design Committee not to be competitive, the OR market was considered not competitive by the system operator (IESO) and thus the OPG is obliged to submit suitable OR offers as detailed in an agreement between OPG and IESO.

If the OR market is also included in the MPMA, it is straightforward to show that the OR price and the *CRQ* must be  $f(q_s^{OR*}, q_s^{E*}, p^F)$  and  $q_s^{OR*}$ , respectively. Then the monopolist's problem becomes:

$$\begin{aligned}
 \underset{M,N}{Max}: & \{g(q^E)q^E - NF - N\bar{q}\} + \{f(q^{OR}, q^E, p^F)q^{OR} - MF\} - \text{rebat}1 - \text{rebat}2 & (4.16) \\
 s.t.: & (1) \quad q^E = N\bar{q} \\
 & (2) \quad q^{OR} = M\bar{q} \\
 & (3) \quad \text{rebat}1 = \begin{cases} 0, & \text{if } g(q^E) \leq g^{-1}(q_s^{E*}) \\ (g(q^E) - g^{-1}(q_s^{E*})) * q_s^{E*}, & \text{if } g(q^E) > g^{-1}(q_s^{E*}) \end{cases} \\
 & (4) \quad \text{rebat}2 = \begin{cases} 0, & \text{if } f(q^{OR}, q_s^{E*}, p^F) \leq f(q_s^{OR*}, q_s^{E*}, p^F) \\ [f(q^{OR}, q_s^{E*}, p^F) - f(q_s^{OR*}, q_s^{E*}, p^F)] * q_s^{OR*}, & \text{if } f(q^{OR}, q_s^{E*}, p^F) > f(q_s^{OR*}, q_s^{E*}, p^F) \end{cases}
 \end{aligned}$$

where *rebat1* is the energy market rebate and *rebat2* the OR market rebate.

Similar to the energy market, the optimal rebate scheme in the OR market would induce the monopolist to bid at the optimal price cap. First, the monopolist would not bid below the price cap, otherwise the marginal profit would be negative. Second, the firm would not bid above the price cap because marginal revenue  $f(q^{OR}, q_s^{E*}, p^F)$  (if the market price is greater than the price cap) is greater than the price cap  $f(q^{OR*}, q_s^{E*}, p^F)$  (resulting from lower production than the *CRQ* for OR), implying that increasing OR production is profitable. As a result, the monopolist will supply OR exactly at the optimal level. As a result, the reliability of the electricity system would be higher if the rebate mechanism is also applied in the OR market.

Therefore, we have the following result 4.

**Result 4**

1. *A socially optimal MPMA requires the CRQ and price cap for energy to be  $q_s^{E^*}$  and  $g(q_s^{E^*})$ , respectively, and the CRQ and price cap for OR to be  $q_s^{OR^*}$  and  $f(q_s^{OR^*}, q_s^{E^*}, p^F)$ , respectively, where  $q_s^{E^*}$  is the minimum value satisfying  $\int_{q_s^{E^*}}^{q_s^{E^*} + \bar{q}} (g(q) - b) dq < F$  and  $q_s^{OR^*}$  is the minimum value satisfying  $\int_{q_s^{OR^*}}^{q_s^{OR^*} + \bar{q}} (f(q, q_s^{E^*}, p^F) - b) dq < F$  ;*
2. *If a socially optimal MPMA is applied to the energy market only, the equilibrium in the energy market would be optimal but OR supply would be too low (suboptimal reliability); in contrast, if the optimal MPMA is applied to both markets, both electricity production and reliability would be optimal.*
3. *If a monopoly generator is capacity constrained and if the MPMA is applied to the energy market only, the monopolist will bid an energy price and an OR price greater than the socially optimal levels. However, the price difference between the energy price and the OR price will be smaller, compared to a non-capacity constrained monopoly.*

**4.9 A Decentralized Market**

The possibility of unit failure, which plays no role in either the social optimum or monopoly, plays a key role in the formation of a competitive equilibrium, because a unit owner must recognize the possibility of a forced outage before he/she enters the market and before he/she starts the generation machine. Thus, the owner's *expected* profit in the

energy market must include the effect of possible unit failure. Most work (Borenstein and Holland, 2003, Chan, Li, and Wu, 2002, etc.) related to reliability issues has ignored the impact of failure possibility on the energy price. Joskow and Tirole (2004) is the only paper that incorporates the probability of production failure within a framework of heterogeneous consumers.

In the long run, a new unit will come into the market if the *expected* revenue can fully cover its fixed cost  $F$  plus variable cost when it produces power or OR. That is, the new investment will be profitable if

$$(1 - p^F)(g(q_c^{E*}) - b) * \bar{q} \geq F \quad (4.17)$$

where  $q_c^{E*}$  is the total system output after the new capacity enters, with subscript  $c$  denoting *competitive*. Note that this is not *perfect* competition in the sense that every entrant will expect that his unit is going to affect the market price once it enters. This is likely true in an electricity market for three reasons. First, market demand is generally predictable, especially in the short and middle term, and thus any potential entrant with sufficient scale would expect a price impact with a new facility. Second, the existing capacity and their approximate costs are publicly known, and thus a potential entrant can assess how he/she can affect the market outcomes. And third, a new unit must enter on a sufficient scale because of economies of scale at the plant level.

In some cases, a new entrant such as a co-generator may not take into account the impact of its entry. A co-generator usually consists of a small plant and thus its fixed cost is small, and its production decision is affected by the co-generated energy e.g. steam.

Then the inequality (4.17) will be easily satisfied, if the fixed cost elasticity of capacity  $\eta_{F\bar{q}} = \frac{\partial F}{\partial \bar{q}} \frac{\bar{q}}{F} > 1$  (i.e. a small increase in capacity will lead to a relatively larger increase in fixed cost). In other words, given the negative sign associated with the failure rate,  $p^F$ , a competitive electricity market is likely to attract small but more reliable units at the margin if a small increase in capacity needs a large increase in fixed cost.

Similarly, inequality (4.17) will more likely be satisfied with a lower  $b$  (i.e. with a more efficient technology) if the fixed cost elasticity of marginal cost,  $\eta_{Fb} = \frac{\partial F}{\partial b} \frac{b}{F} > -1$ , given the fact that fixed cost is negatively related to variable cost in the electricity industry (i.e.  $\eta_{Fb} < 0$ ).

In addition, inequality (4.17) is more likely satisfied, the smaller is the average fixed cost,  $F/\bar{q}$ . That is, the possibility of entry is negatively related to fixed cost. A special case is where  $F$  is equal to zero, where the outage rate has no impact on the investment decision: once the expected price is higher than marginal cost, a new plant should be built.

Solving (4.17) yields the competitive energy supply  $q_c^{E*} = g^{-1}\left(\frac{F}{\bar{q}(1-p^F)} + b\right)$  and the corresponding energy price  $\frac{F}{\bar{q}(1-p^F)} + b$ . It is apparent that neither marginal cost nor average cost pricing is sustainable. The equilibrium energy price is positively related to the unit failure rate: a higher probability of failure requires a higher price to compensate for the investment cost. This result highlights a common problem in electricity market

simulations with the so-called “net revenue analysis”<sup>39</sup>. This approach assumes that a unit with a specific heat rate may generate electricity as long as the market price is greater than its heat rate. Then total revenue for this unit for a certain period is calculated. If the revenue covers levelized fixed cost plus variable cost, the unit is said to break even. This approach, however, could significantly overstate the true revenue for a fossil fuelled generator which in general has a high failure rate. As a consequence, the comparison of *net revenue* may undermine the importance of a high electricity price. A more meaningful calculation should incorporate the probability of production failure and unavailability for regular maintenance although the highest priced periods might be avoided.

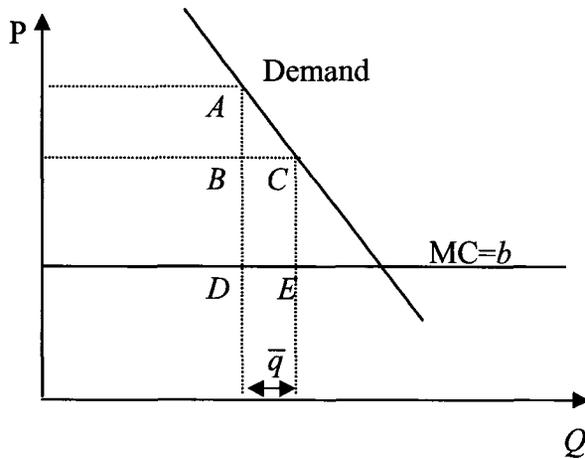
Compared to the socially optimal energy supply and price, the competitive output must be lower, and the competitive price must be higher. The difference is illustrated in Figure 4.3. The social optimum is obtained by comparing the fixed cost  $F$  to the area between the downward-sloped demand curve and the horizontal line  $b$ , while the competitive equilibrium is obtained by comparing adjusted fixed cost  $F/(1 - p^F)$  with the area between the market price and the horizontal line  $b$ . In particular, to determine whether or not a new generating unit should be constructed, the social optimum will compare the area  $ADECA$  with fixed cost  $F$ : if the area  $ADECA$  is greater than  $F$ , the unit should be added into the market; otherwise no new capacity should be built. In contrast, a competitive firm will compare the area  $BDECB$  with  $F/(1 - p^F)$ . If the area  $BDECB$  is

---

<sup>39</sup> *Net revenue* usually equals the sum of the positive hourly difference between market price and marginal cost in one year. The ISO's then compare *net revenue* to fixed cost: if *net revenue* is greater than the fixed cost, the unit is said to be viable.

greater than  $F/(1-p^F)$ , the firm will enter the market or invest in a new facility; otherwise it will stay out or make no new investment. Given that the area  $ADECA$  is greater than the area  $BDECB$  and  $F$  is smaller than  $F/(1-p^F)$ , the competitive market supply must be lower than the social optimum, and thus the competitive price is higher.

**Figure 4.3: The Difference between the Social Optimum and Competitive Equilibrium**



Similarly, a new investment will be profitable in the OR market if

$$f(q_c^{OR*}, q_c^{E*}, p^F) * \bar{q} \geq F \quad (4.18)$$

where  $q_c^{E*}$  is the competitive energy output obtained from (4.17), and  $q_c^{OR*}$  is the total OR supply after the new capacity enters. Note that the inequality doesn't contain the multiplier  $(1-p^F)$  because a unit generally has no outage when it supplies OR. Solving the inequality (4.18) yields the competitive OR supply  $q_c^{OR*} = q_c^{OR*}(f^{-1}(F/\bar{q}), q_c^{E*}, p^F)$  and OR price  $F/\bar{q}$ . The long run price difference between the energy price and the OR price is thus  $\frac{Fp^F}{\bar{q}(1-p^F)} + b$ , which is greater than the difference in marginal cost. If the generator never fails, i.e.,  $p^F = 0$ , the difference shrinks to  $b$ . Borenstein and Holland

(2003) assume the demand for OR is a fixed proportion of the demand for energy, and find that the difference between the energy and OR prices is simply the difference in marginal cost. Thus their model which does not take the forced outage rate into account is a special case of a more general model with an outage rate.

One can see that the cost of constructing a generating unit is fully covered in either the energy market or the OR market, and thus a firm is indifferent between investing or bidding in either market.

If the OR requirement is deterministically established and less than the competitive OR supply based on the probabilistic approach, the equilibrium energy and OR prices will be the same as above. Yet the OR supply and system reliability are lower. On the other hand, if the deterministic OR requirement is greater than the competitive OR supply, the equilibrium energy and OR prices and the energy supply will be unchanged in the long run, but the OR supply is too large and thus reliability is too high.

In the short run, if the total installed capacity is less than total demand (energy plus OR), for example, during peak periods, a price taker will have to decide which market to bid in. A firm should arbitrage the profit difference between the energy and OR markets, such that in equilibrium the *expected* profit in either market will be exactly the same, that is,  $f(q_{cs}^{OR*}, q_{cs}^{E*}, p^F) * \bar{q} = (1 - p^F)(g(q_{cs}^{E*}) - b) * \bar{q}$ , where subscript *cs* denotes *competitive- and-capacity-constrained*. Rearranging the equilibrium condition yields the price difference  $g(q_{cs}^{E*}) - f(q_{cs}^{OR*}, q_{cs}^{E*}, p^F) = b + p^F (g(q_{cs}^{E*}) - b)$ . Therefore the difference

between the energy and OR prices will be greater than the marginal cost of producing energy (given that  $g(q_{cs}^{E*}) > b$ ), which could be greater or smaller than the long-run price difference  $\frac{Fp^F}{\bar{q}(1-p^F)} + b$ . The price difference also depends on the extent of the shortage, which is implicitly measured by  $(g(q_{cs}^{E*}) - b)$  -- the larger the shortage, the greater the price gap.

Given the price difference and the capacity limit, the two unknown variables  $q_{cs}^{E*}$  and  $q_{cs}^{OR*}$  can be solved. Specifically, the following system can be solved for the equilibrium energy and OR output:

$$g(q_{cs}^{E*}) = f(q_{cs}^{OR*}, q_{cs}^{E*}, p^F) / (1 - p^F) + b \quad (4.19 a)$$

$$q_{cs}^{E*} + q_{cs}^{OR*} = L\bar{q} \quad (4.19 b)$$

where  $L$  is the total number of available units.

From (4.19a), it can be shown that  $\partial g(q_{cs}^{E*}) / \partial p^F > 0$ . That is, a higher failure rate will lead to a higher energy price. On the other hand, the energy price also depends on the OR demand curve, and thus a simple model that assumes the demand for energy to be the sum of the demand for energy and for OR is too simple in a joint optimization framework. These two results illustrate the over-simplicity of some simulation techniques that have ignored unit malfunctions, for example, Mansur (2001), and Bushnell and Saravia (2002). Harvey and Hogan (2002) also point out other problems with those simulations, for example, data quality and/or model specification errors.

In the short run, if total demand (energy plus OR) is lower than total installed capacity (i.e. there is excess capacity), given that all fixed costs are sunk, the market equilibrium will be very simple: every firm will bid at marginal cost or avoidable cost and unit outage is not a problem. The intuition is that if a firm tries to incorporate unit failure into its bidding strategy, it will never be in the money. Thus, the equilibrium price is  $b$  for energy and  $0$  for OR, and the respective supply is  $g^{-1}(b)$  for energy and  $q^{OR}(f^{-1}(0), g^{-1}(b), p^F)$  for OR. Therefore, when there is excess capacity in the marketplace, the simulation techniques widely used in the literature would approximate the market short run cost, but may lead to long run resource inadequacy. Doorman and Nygreen (2002) construct a simulation model based on short-run cost and unit constraints and confirm that for most hours when supply is sufficient the market equilibrium price will be around marginal cost. When demand is high, the energy price will be well above marginal cost and the OR price closely follows the energy price (that is the OR price is the opportunity cost of supplying energy)

The deterministic approach to establishing the OR requirement could, in the short term, lead to no equilibrium at all when there is a shortage in capacity. Assume there is only one unit short. If the shortage is in the OR market, then the OR price will be infinite or a predetermined price cap which is generally high, if any, which will attract a unit away from the energy market. Once the unit enters the OR market, however, the OR price will immediately drop to zero. At the same time, with the unit leaving the energy market, the energy price becomes higher. Then the energy market will become more attractive, and a unit should move away from the OR market to the energy market, which lowers the

energy price, but the OR price will now be infinite or the price cap again. Once again, the OR market becomes more attractive. As a result, there is no short-run equilibrium: the market would be cleared only by chance.

To respond to a shortage condition, the general practice by a system operator is to lower the operating reserve requirement. This brings down the market price to the competitive level without seriously affecting reliability. However, this reduction in operating reserve requirement distorts the market signal and is a disincentive to investment (MSP, October 2002, December 2005, Joskow and Tirole, 2004).

This non-existence of a competitive equilibrium could lead to frequent price spikes in a dynamic context in either market, which signals resource scarcity in the electricity market, and thus should attract more investment. As was mentioned above, if the deterministic OR requirement is greater than the optimal level, the price signal may attract too much investment. In contrast, if the deterministic OR requirement is less than the optimal level, there may be insufficient investment. Joskow and Tirole (2004) find that ‘under certain contingencies the market price, and the associated scarcity rents available to support investment in generating capacity, are extremely sensitive to small mistakes or discretionary actions by the system operator’ (page 47).

An extremely low OR price cap is a very interesting topic because, as mentioned before, some ISO’s do impose a very low OR price cap in certain areas where the ISO determines the OR market is not competitive. For example, in the early years of the

California wholesale market, the OR price was regulated at a very low level. When the demand in the area is low, the price cap won't be binding. However, if supply becomes tight frequently, the OR price cap could have a large impact on the energy market as well as on investment decisions.

Because the OR price cap is low, bidding in the OR market would be un-attractive and thus a firm will have a strong incentive to bid its capacity into the energy market first. The outcome will likely be a relatively lower energy price than it would prevail if an OR price cap were not in place. If there is sufficient capacity in the marketplace, an equilibrium in the energy market will be reached with the energy price equal to the marginal cost  $b$ . Any remaining capacity goes to the OR market. Some market participants may not offer into the OR market at all. This is exactly what happened in California during much of 1998, when there was "bid insufficiency" in the OR market as a result of the regulated low OR price (Hirst, 2000). Because the short-run prices don't signal resource scarcity when it actually exists, there will be no new investment and the system is underinvested.

***Result 5 (competitive equilibrium)***

*In the long run,*

1. *The competitive equilibrium energy and OR prices are generally higher than their respective socially optimal prices or marginal cost pricing: the competitive energy price and supply are  $\frac{F}{\bar{q}(1-p^F)} + b$  and  $q_c^{E*} = g^{-1}(\frac{F}{\bar{q}(1-p^F)} + b)$ , respectively,*

*and the competitive OR price and supply are*

*$F/\bar{q}$  and  $q_c^{OR*} = q_c^{OR*}(f^{-1}(F/\bar{q}), q_c^{E*}, p^F)$ , respectively;*

2. *The possibility of machine failure will affect offer strategies and investment decisions as well; and*
3. *The deterministic approach for establishing the OR requirement may lead to either too high or too low investment.*

*In the short run,*

4. *when there is a supply shortage, both energy and OR prices will be higher, and the difference between the energy price and the OR price is greater than the difference between their respective marginal costs;*
5. *when there is a supply shortage, the deterministic OR requirement could lead to no equilibrium at all; To clear a tight market, a system operator may lower the requirement for operating reserve, which has potential implications for investment;*
6. *simulation of system marginal cost is valid only when the supply is sufficient and thus competition is intense; and*
7. *a low OR price cap will lower the energy price when there is a supply shortage (contrary to the monopoly case).*

#### 4.10 Capacity Allocation in Shortage Conditions: Monopoly vs. Competition

A question of interest is how the producers allocate their limited resources in different market structures during shortage periods.

We repeat the FOC's for a capacity-constrained monopoly:

$$g(q^E) + q^E \frac{dg(q^E)}{dN} - \frac{F}{q} - b + q^{OR} \frac{\partial f(q^{OR}, q^E, p^F)}{\partial N} - \lambda = 0 \quad (4.10 a)$$

$$f(q^{OR}, q^E, p^F) + q^{OR} \frac{\partial f(q^{OR}, q^E, p^F)}{\partial M} - \frac{F}{q} - \lambda = 0 \quad (4.10 b)$$

$$q^E + q^{OR} = L\bar{q} \quad (4.10 c)$$

Combining the first two FOCs yields:

$$g(q^E) = f(q^{OR}, q^E, p^F) + q^{OR} \left( \frac{\partial f(q^{OR}, q^E, p^F)}{\partial M} - \frac{\partial f(q^{OR}, q^E, p^F)}{\partial N} \right) - q^E \frac{dg(q^E)}{dN} + b \quad (4.10 a'')$$

$$q^E + q^{OR} = L\bar{q} \quad (4.10 c'')$$

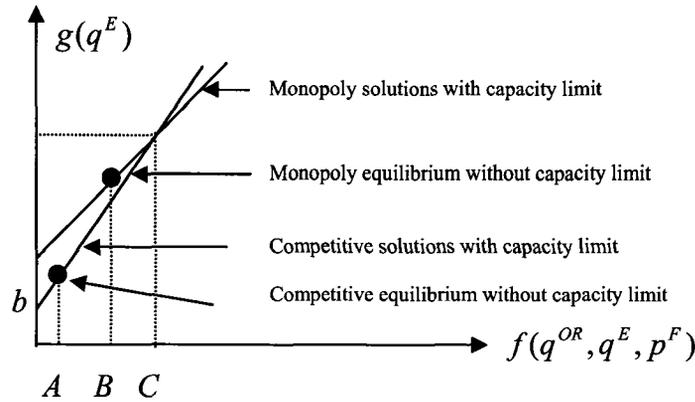
In contrast, the FOC's for a competitive equilibrium are:

$$g(q_{cs}^{E*}) = f(q_{cs}^{OR*}, q_{cs}^{E*}, p^F) / (1 - p^F) + b \quad (4.19 a)$$

$$q_{cs}^{E*} + q_{cs}^{OR*} = L\bar{q} \quad (4.19 b)$$

Thus, if equation (4.10a'') is equal to (4.19a), the monopolist will allocate its limited resources in the same way as does a competitive industry. Equation (4.10a'') and (4.19a) are illustrated in Figure 4.4 with (4.19a) having a steeper slope as a result of a positive  $p^F$ . Note that the line corresponding to equation (4.10a'') can shift up or down depending on the total available capacity, which results from the non-linear relationship between the price and quantity demanded. The intersection point indicates that the outcome of the competitive equilibrium can be the same as the monopoly outcome.

**Figure 4.4: Equilibrium Energy and OR Prices**



In particular, the energy price difference between competition and monopoly can be written as:

$$g(q_{cs}^{E*}) - g(q^E) = \frac{f(q_{cs}^{OR}, q_{cs}^{E*}, p^F)}{1-p^F} - f(q^{OR}, q^E, p^F) - q^{OR} \left( \frac{\partial f(q^{OR}, q^E, p^F)}{\partial M} - \frac{\partial f(q^{OR}, q^E, p^F)}{\partial N} \right) + q^E \frac{dg(q^E)}{dN} \quad (4.20)$$

If equation (4.20) is equal to zero, both the monopoly and competitive markets allocate a given quantity of resources in the same way, and the equilibrium prices under monopoly and competition are the same. This implies that the output ratio is

$\frac{q^{OR}}{q^E} = \left( \frac{dg(q^E)}{dN} + \frac{p^F f(\bullet)}{(1-p^F)q^E} \right) / \left( \frac{\partial f(\bullet)}{\partial M} - \frac{\partial f(\bullet)}{\partial N} \right)$ , which can be further modified as follows:

$$\frac{f(\bullet)}{g(q^E)} = \frac{\eta_{gN}}{\eta_{fM} - M\eta_{fN} / [N - p^F / \{\bar{q}(1-p^F)\}]} \quad (4.21)$$

where  $\eta_{gN}$  is the inverse demand elasticity of energy,  $\eta_{fM}$  is the inverse demand elasticity of OR, and  $\eta_{fN}$  is the cross-elasticity of energy demand with respect to the OR price. Therefore, if the optimal bidding strategy for the monopolist is to bid the relative price of OR and energy equal to the ratio of the inverse demand elasticity of energy to the

*adjusted* inverse demand elasticity of OR, the monopoly resource allocation would be the same as under competition.

Figure 4.4 above illustrates this comparison between the monopoly equilibrium and the competitive equilibrium. The monopoly solution must be on the right hand side of  $B$  (inclusive), where  $B$  is the equilibrium for monopoly without capacity constraint. All competitive solutions must be on the right side of  $A$  (inclusive), where  $A$  is capacity-unconstrained competitive equilibrium. Between  $B$  and  $C$ , both monopoly and competition are capacity constrained but with the monopoly energy price is higher than the competitive price. The sets beyond point  $C$  are equilibria where the competitive energy price is higher than the monopoly energy price when the unit outage rate is very high.

In fact, the demand for either energy or OR is not very elastic in the short term, that is, the elasticity  $\eta_{gN}$  and  $\eta_{fM}$  are both very large, but the cross elasticity  $\eta_{fN}$  is relatively small. Then the *adjustment* term can be dropped from the denominator of equation (4.21), and the equation becomes  $f(\bullet)\eta_{fM} = g(q^E)\eta_{gN}$ . In this case, the monopoly allocation would be exactly the same as under competition if the OR price divided by the elasticity of OR demand is equal to the energy price divided by the elasticity of energy demand.

Similarly, the competitive energy price would be greater than monopoly energy price (implying competition leads to fewer resources in the energy market)

if  $f(\bullet)\eta_{fM} > g(q^E)\eta_{gN}$ . For this condition to hold,  $\eta_{fM}$  must be very high given that the OR price  $f(\bullet)$  is usually lower than energy price  $g(q^E)$  even during shortage, or near-shortages.

How can  $\eta_{fM}$  be very high? One obvious answer is if the unit outage rate  $p^F$  is relatively large. Remember that the competitive equilibrium requires the expected profit in the energy market to be equal to the expected profit in the OR market, thus a higher forced outage rate needs a higher energy price to cover the fixed cost. In contrast, the forced outage rate has no impact on the monopolist's profit in the energy market. As a consequence, the competitive energy price could be higher than the monopoly energy price and, correspondingly, the competitive OR price is lower than the monopoly OR price.

**Result 6** (*monopoly vs. competition with capacity constraint*)

1. *If the unit outage rate is very high, the competitive energy output would be lower than the monopoly energy output, and consequently the competitive OR output is higher than the monopoly OR output; and*
2. *If the unit outage rate is low, the competitive energy output would be higher than the monopoly output and the competitive OR output is lower than the monopoly OR output.*

#### 4.11 Policies Leading to the Social Optimum

As was shown above, there are two key factors leading to a deviation of the competitive equilibrium from the social optimum: fixed cost and the possibility of machine malfunction. Unit malfunction is not a problem in the social welfare maximization because any outage will be internalized by the stand-by OR and the OR demand curve has already captured all possibilities of system failure, while it is a problem for a competitive firm because it makes no revenue if its machine fails to produce energy.

Traditionally, marginal cost pricing requires a lump sum transfer if there are economies of scale, or fully distributed cost pricing, or Ramsey pricing is implemented to induce the second best solution if a transfer is not allowed. The unit failure problem is ignored in most regulation literature. One possible reason for this is that firms can build up inventories so that machine failures have little impact on sales. However, this is not true in the electricity industry, because electricity is not economically stored or is essentially impossible to store with current technology.

In the OR market, however, there is no outage problem. Furthermore, the OR market is a derivative market, thus once the energy market is at its optimum, the OR market must be automatically at its optimum if there are no supply shortages and no regulatory restrictions in the OR market (such as a price cap). Therefore, we simply need to design optimal policies for the energy market, leaving the OR market to clear automatically.

Public policy has long been focusing on capacity subsidies or payments to induce sufficient investment to deal with the “missing money” problem - due to price caps in energy markets or other practices such as the system operator’s control actions that depress the market price (Hogan, 2005, Joskow, 2005, Borenstein and Holland, 2003). For example, the northeastern ISO’s in the United States have implemented capacity markets, to deal with the “missing money” problem. In contrast, there is no such capacity market in Ontario, Australia, and Alberta. Ontario is currently considering implementing a capacity market.

Many analysts (e.g. Hirst and Hadley, 1999, Oren, 2005) believe that a capacity market is a fictional market because there is no intrinsic value, and is not needed for inducing new investment (as demonstrated in Alberta and Australia). Others argue that the capacity market is necessary because of either political constraints (e.g. Besser, Farr, and Tierney, 2002), price caps (Hogan, 2005), the effect of market power mechanisms on scarcity rent (Joskow, 2005), efficiency improvements based on simulation (e.g. Hobbs, Inon, and Stoft, 2001), or insufficient payment for reliability (e.g. Bowring and Gramlich, 2000, Joskow, 2005).

Given the existence of unit outages and the existence of above constraints, the capacity market may to some extent help improve the investment environment. Recall that we have two important results: (i) that the long run competitive investment will be lower than the social optimum, and (ii) that competitive investment is negatively related to average fixed cost. The former says that the social optimum requires more investment

than the competitive equilibrium can provide, and the latter says that more investment can be induced by lowering the fixed cost. Thus if we can lower the average fixed cost through some mechanism, we can improve the long run investment efficiency. This is consistent with Joskow (2005) who argues that due to the probability distribution of peak demand realization and generation unit availability, a capacity market can serve as a ‘safety valve’ to produce adequate investment when other constraints limit the ability of the shortage price to reflect the value of lost load.

Because capacity subsidies are not possible due to political sensitivity<sup>40</sup>, a capacity market may be an alternative. With a capacity market, a part of the fixed cost is covered, and thus the condition (4.17) is more likely to be satisfied.

***Result 7** Capacity subsidies or capacity markets can to some extent help solve the unit outage problem, can improve the investment environment, and could be welfare-improving.*

#### **4.12 Conclusions**

The reliability issue is largely ignored by most economists possibly because of the low price of operating reserve, the limited significance of market power in the OR market and the number of engineering complications that are difficult to model. It is true, under normal conditions where there is no significant shortage of capacity, that reliability is not

---

<sup>40</sup> Another way to reimburse the fixed cost is to lower the tax burden of generators, such as the tax break for renewable power generation in Ontario.

at risk and the OR price is usually very low<sup>41</sup>. Given that the demand for OR is generally small (usually about 10 percent of energy demand), its market value is considerably less than the value of the energy market. Furthermore, even if the market faces shortages of supply which may lead to no equilibrium, a system operator can take out-of-market actions to lower demand for OR<sup>42</sup>. As a result, both the energy and OR markets can clear all the time.

As was demonstrated above, however, ignoring the role of the OR market in joint optimization leads to significant flaws in assessing market power. The coexistence of the energy and OR markets adds more caveats to the interpretation of market simulation results, besides the sensitivity problem raised by Harvey and Hogan (2002).

This chapter also analyzes the joint optimization of energy and OR markets when the demand for OR is determined by either a statistical approach (a possible approach in the future) or a deterministic approach. It shows that the market can always be cleared when the demand for OR is statistically determined which reduces the need for the system operator to take control actions that may depress market prices, while there could be no equilibrium when demand for OR is deterministically established and when market is tight. It also shows that the two reliability standards set by NERC are inconsistent. That

---

<sup>41</sup> If the markets are sequentially cleared like the early stage of California's market, the OR price could be much higher than energy price. The design flaw in California was corrected later.

<sup>42</sup> For example, in Ontario, the IESO can lower OR requirements by the shortage amount so that the market can clear all the time. In addition to lowering the OR requirement, a system operator could reduce the voltage requirement in an area where demand is higher than supply. The voltage reduction could, however, damage the voltage-sensitive appliances.

is, the criterion of one-day-in-ten-years might be inconsistent with the n-1 standard or capacity requirement.

Finally, an important theoretical implication is that capacity subsidies or a capacity market can be welfare improving.

## References

1. Baron, D. and R. Myerson (1982) "Regulating a Monopolist with Unknown Costs", *Econometrica* 50:911-930.
2. Beenstock, M., E. Goldin, and Y. Haitovsky (1998) "Response bias in a conjoint analysis of power outages", *Energy Economics* 20:135-156.
3. Besser, J., J. Farr and S. Tierney (2002) "The Political Economics of Long Term Generation Adequacy: Why an ICAP Mechanism is Needed as a Part of Standard Market Design?", *Electricity Journal* August/September:53-62.
4. Billinton, R. and R.N. Allen (1996) *Reliability Evaluation of Power Systems*, Plenum Press, New York.
5. Borenstein, S. and S. Holland (2003) "Investment Efficiency in Competitive Electricity Markets With and Without Time-Varying Retail Prices", University of California Energy Institute, Working Paper PWP-106.
6. Bowring, J. and R. Gramlich (2000) "The Role of Capacity Obligations in a Restructured Pennsylvania-New Jersey-Maryland Electricity Market", *Electricity Journal* 13(November):57-67.
7. Bushnell, J. and C. Saravia (2002) "An Empirical Assessment of Competitiveness of the New England Electricity Market", University of California Energy Institute, Working Paper PWP-101.
8. Carlton, D.W. (1977) "Peak Load Pricing with Stochastic Demand", *American Economic Review* 67(5):1006-1010.
9. Chan, C. Y., Y. X. Li and F. F. Wu (2002) "Fundamentals of Operating Reserve Procurement Using Decentralized Approach with Insurance Theory Application", *Electric Power Systems Research* 63:229-238.
10. Cramton, P. and S. Stoft (2006) "The Convergence of Market Design for Adequate Generating Capacity", Manuscript, April 25, 2006.
11. Crew, M. and P. Kleindorfer (1976) "Peak Load Pricing with a Diverse Technology", *Bell Journal of Economics* 7(Spring):232-241.
12. ----- (1978) "Reliability and Public Utility Pricing" *American Economic Review* 68(March):31-40
13. Doorman, G. and B. Nygreen (2002) "An integrated model for market pricing of energy and ancillary services", *Electric Power Systems Research* 61:169-177.

14. Harvey, S. and W. Hogan (2002) “Market Power and Market Simulations”, LECG, LLC., and Harvard University, July 16, available at: [http://ksghome.harvard.edu/~whogan/H-H\\_Market\\_Power&Simulations\\_071602.pdf](http://ksghome.harvard.edu/~whogan/H-H_Market_Power&Simulations_071602.pdf)
15. Hirst, E. (2000) “Maximizing Generator Profits Across Energy And Ancillary Service Markets”, *Electricity Journal* 13(June):61-69.
16. Hirst, E. and S. Hadley (1999) “Generation Adequacy: Who Decides?”, *Electricity Journal* 12(October):11-21.
17. Hirst, E. and B. Kirby (1997a) “Operating Reserve and Bulk-power Reliability”, *Energy Economics* 23(11):949-959.
18. Hirst, E. and B. Kirby (1997b) “Ancillary Service Details: Operating Reserves”, U.S. Department of Energy, Contract No. DE-AC05-96OR22464.
19. Hirst, E. and B. Kirby (1998) “Technical and Market Issues for Operating Reserves”, *Consulting in Electric-Industry Restructuring*, Oak Ridge, Tennessee.
20. Kirby, B. and E. Hirst (2003) “Allocating the Cost of Contingency Reserves”, *Electricity Journal* 16(December):39-47.
21. Hogan, W. (2005) “On an ‘Energy Only’ Electricity Market Design for Resource Adequacy”, Centre for Business and Government, John F. Kennedy School of Government, Harvard University, available at: [http://ksghome.harvard.edu/~whogan/Hogan\\_Energy\\_Only\\_092305.pdf](http://ksghome.harvard.edu/~whogan/Hogan_Energy_Only_092305.pdf)
22. Hobbs, B., J. Inon and S. Stoft (2001) “Installed Capacity requirements and Price Cap: Oil on the Water, or Fuel on the Fire?”, *Electricity Journal* 14(July):23-34.
23. Joskow, P. (2005) *Why Capacity Obligations and Capacity Markets?*, available at: [http://econ-www.mit.edu/faculty/index.htm?prof\\_id=pjoskow&type=paper](http://econ-www.mit.edu/faculty/index.htm?prof_id=pjoskow&type=paper)
24. Joskow, P. and J. Tirole (2004) “Reliability and Competitive Electricity Markets”, University of California Energy Institute, Working Paper PWP-129.
25. Kariuki, K. and R. Allan (1996) “Evaluation of Reliability Worth and Value of Lost Load”, *IEEE Proceedings—Generation, Transmission, and Distribution* 143:171-180.
26. Laffont, J-J. and J. Tirole (1994) *A Theory of Incentives in Procurement and Regulation*, The MIT Press, Cambridge, Massachusetts.

27. Mansur, E. (2001) "Pricing Behaviour in the Initial Summer of the Restructured PJM Wholesale Electricity Market", University of California Energy Institute, Working Paper PWP-083.
28. Market Surveillance Panel, *Market Monitoring Report*, October 2002, December 2005, [www.oeb.gov.on.ca](http://www.oeb.gov.on.ca).
29. Oren, S. (2005) "Generation Adequacy Via Call Options: Safe Passage to the Promised Land", University of California Energy Institute, Working Paper PWP-016.
30. Siddiqi, S. N. and M. L. Baughman (1995) "Reliability Differentiated Pricing of Spinning Reserve", *IEEE Transactions on Power Systems* 10(3):1211-1218.
31. Stoft, S. (2002) *Power System Economics: Designing Markets for Electricity*, IEEE Press, N.J.
32. Sweeting, A. (2004) "Market Power in the England and Wales Wholesale Electricity Market 1995~2000", Center for Energy and Environmental Policy Research Working Paper 0413, Massachusetts Institute of Technology
33. Telson, M. L. (1975) "The Economics of Alternative Levels of reliability for Electric Power Generation Systems", *Bell Journal of Economics* 6(2):679-694.
34. Wachter, G. and R. Billinton (1989) "Customer Cost of Electric Service Interruptions", *Proceedings of the IEEE* 77:919-930.
35. Willis, K. and G. Garrod (1997) "Electricity Supply Reliability – Estimating the Value of Lost Load," *Energy Policy* 25:97-103.

## Chapter 5 Conclusions

The intent of this thesis is, first, to analyze the impact of the Market Power Mitigation Agreement (MPMA) signed between the Ontario government and Ontario Power Generation Inc. (OPG) on the offer strategy of OPG, and, second, to examine the effect of electric power system reliability requirements on the market for electric power (the energy market). Chapters 2 and 3 focus mainly on the impact of the price cap and rebate mechanism of the MPMA which required OPG to pay back a certain portion of its revenue when the quantity-weighted market price was above \$38/MWh. The quantity that was used to calculate the average price was 90% of simulated output of OPG based on the assumption that all other firms bid competitively and that OPG bid at its average cost. Chapter 4 constructs a linkage between the demand for electricity and system reliability by establishing an endogenous demand function for operating reserve (OR) based on a statistical model.

The price cap and rebate mechanism under the MPMA was unique in the sense that no other electricity markets or other commodity markets have adopted this approach to mitigating market power of dominant producers. There were also other approaches that might have led to better market outcomes, such as dissolving the OPG into several competing companies. The reason why the Design Committee chose this mitigation mechanism was to serve as an interim regime until OPG's generation asset were divested.

After four-years, the MPMA was terminated. The price cap and rebate mechanism currently plays a more limited role in mitigating the market power of OPG. The new Electricity Restructuring Act 2004 (Bill 100) of Ontario significantly changed the course of regulation and most electricity production by OPG is presently subject to a fixed price<sup>43</sup>. Although OPG is still the largest producer who has the most influence on the market price, its incentive to exercise market power is mitigated in a variety of ways. As a result, the current market structure is a hybrid market in which the government is deeply involved.

This thesis mainly discusses the impact of the MPMA on OPG's strategy before the provincial new restructuring effort in 2004 and compares it with the socially optimal price cap regime. In Chapter 2, we construct a one period model which explores the consequences of the MPMA on a monopolist's bidding strategy and derives the socially optimal price cap and rebate mechanism. In Chapter 3, we extend the model to two periods and incorporate hydroelectric generation that can be used in either period. In Chapter 4, we build a linkage between reliability requirements and energy supply/demand, and explore its consequences.

In Chapter 2, we derived the following results.

1. *The price cap mechanism in the MPMA is welfare-improving as long as it is lower than the unrestricted monopoly price and the CRQ is a non-zero number, regardless of the efficiency type of the monopolist.*

---

<sup>43</sup> The early movers (those built plants before Bill 100) also signed a contract with the government as a result of new intervention by government.

2. *The MPMA is second best only if production is below the level at which average cost reaches its minimum. If output is such that average cost is above its minimum point, the monopolist's profit-maximizing bid is higher than the price cap.*
3. *The Price Spike Adjustment (PSA) doesn't increase the monopolist's incentive to bid above the cap.*
4. *As long as  $CRQ > \frac{A-\theta}{B}$ , social optimality can be achieved by simply setting the optimum price cap, implying that there are multiple CRQs that could lead to the social optimum.*
5. *If the contract is designed without taking into account demand uncertainty, the monopolist will be worse off under a negative demand shock while better off under a positive demand shock. Consumer surplus is never maximized, that is, there are some consumers whose consumption will be rationed.*
6. *There exists no optimal rebate mechanism under demand uncertainty. In fact, the monopoly firm will produce less than the socially optimal level which maximizes combined consumer surplus and producer surplus.*
7. *Under demand uncertainty, the second best MPMA requires a higher CRQ than the expected break even output, and thus a lower price cap. The soft price cap and rebate mechanism is superior to a fixed price cap because it allows risk sharing between consumers and the regulated firm.*
8. *If the regulatory regime is designed without taking into account supply uncertainty, the equilibrium won't be the social optimum and the price cap will serve as the lower bound of market prices. If the supply shock is positive,*

*e.g. an increase in fuel cost, the market price will be the cap, the monopolist experiences a loss and produces too much, while if the supply shock is negative, the market price will be higher than the cap, the firm makes a positive profit and produces too little.*

9. *If an outage reduces capacity below the CRQ, the monopolist may be better off to import electricity to mitigate the price impact (eventually lowering the rebate to consumers), even if the import price is higher than the domestic price. However it may never import if the external price is high, either because the price is too high or because the Price Spike Adjustment insulates its revenue from the impact of a high price. The firm will suffer losses whether or not it imports electricity.*

The main contribution in Chapter 3 is the introduction of an energy-limited resource, in particular, a hydroelectric resource. Conclusions reached in this chapter may be extended to other situations where the power supply is limited by fuel inputs or environmental regulations. For example, a generating unit's output may be limited either by the shortage of fuel supplied due to a transportation interruption or by environmental regulations such as emission limits.

We have the following conclusions in this chapter.

1. *There are multiple policy choices to induce the monopolist to bid at the pre-specified price cap.*

2. *The MPMA price cap is welfare improving as long as the price cap is lower than the CRQ-weighted unconstrained monopoly price  $AP_m$ . Peak consumers benefit more from the imposition of the price cap.*
3. *Under the MPMA price cap, a monopoly generator would shift hydro production to the peak period.*
4. *The first best price cap is unique. The first best CRQ's are multiple, as long as they are greater than or equal to the minimum optimal CRQ's.*
5. *If a positive demand shock (i.e. demand increases unexpectedly) happens in the first period, the average price will be higher than the price cap. The second period consumer will share the cost of the demand shock in the first period. In this situation, the monopolist makes positive profit. In contrast, in the event of a negative demand shock, the equilibrium price would be the same for both periods, and the monopolist makes a zero profit. An unconstrained monopolist will respond to the demand shock in the same way as a constrained monopolist [i.e. increasing production by the same amount] if the demand shock is positive, and will respond to the shock more efficiently if the shock is negative.*
6. *If the outage happens in the first period, the monopolist will not import any electricity, will always bid above the price cap, shift hydro production from the second period to the first period, and increase thermal production in the second period.*
7. *As an alternative to the MPMA, marginal cost pricing in general won't lead to a social optimum; the water resource will be inefficiently used. To achieve*

*the social optimum through marginal cost pricing, the monopolist could be dissolved into one firm with thermal capacity only and subject to the marginal cost pricing, and many competing hydro firms.*

In Chapter 4, we try to bridge the gap between the system reliability level defined using engineering criteria and the economic (market) demand for operating reserve (OR). We find that:

1. *Without a capacity constraint, the monopoly equilibrium will have the following characteristics:*
  - (A) *The energy price is not necessarily greater than the OR price although the cost of providing OR is lower than the cost of providing energy;*
  - (B) *The monopolist will produce more electricity so as to increase the demand for operating reserve when a probabilistic OR pricing approach is applied; and*
  - (C) *The deterministic approach to calculating the OR requirement may lead to either higher or lower reliability than the optimum reliability level; and the deterministic approach will lead to higher energy and OR prices, and lower energy production;*
2. *When a monopolist's production capacity is limited,*
  - (A) *when the OR requirement is probabilistically determined, the monopolist will increase both the energy and the OR price by the same amount relative to prices when there is no capacity constraint; or*

- (B) *when the OR requirement is deterministically established, a monopolist will allocate generation first in the most profitable market with the remaining resources in the other market.*
3. *The socially optimal price is not necessarily equal to marginal cost in the long run. The reason for this is that the increase in consumer surplus may be totally offset by the increase in fixed cost when a marginal unit is to be added to the system.*
4. *If a socially optimal MPMA is applied to the energy market only, the equilibrium in the energy market would be optimal, but the OR supply would be too low;*
5. *In a competitive market and in the long run,*
- (A) *the competitive equilibrium energy and OR prices are generally higher than their respective socially optimal prices or marginal cost pricing.*
- (B) *the possibility of machine failure will affect offer strategies and investment decisions; and*
- (C) *the deterministic approach to establishing the OR requirement may lead to either too high or too low investment.*
6. *In a competitive market and in the short run,*
- (A) *when there is a capacity limit and the OR requirement is statistically determined, both the energy and OR prices will be higher, but the difference between the energy price and the OR price is greater than the difference between their respective marginal costs;*
- (B) *when there is a capacity limit, the deterministic OR requirement could lead to no equilibrium at all ; to clear a tight market, a system*

*operator may reduce the requirement for operating reserve, which has an adverse impact on new investment;*

*(C) simulation of system marginal cost is valid only when the supply is sufficient and thus competition is intense; and*

*(D) a low OR price cap will reduce the energy price when there is a supply shortage (contrary to the monopoly case).*

*7. Capacity subsidies or capacity markets can to some extent help solve the unit outage problem, can improve the investment environment, and may be welfare-improving.*

## Glossary

### **Dispatchable Load**

Customer demand that is responsive to price-based dispatch directions from the System Operator.

### **Load Shedding**

The process of deliberately removing (either manually or automatically) pre-selected customer demand from the power system in response to an abnormal condition to maintain the integrity of the system and minimize overall customer outages.

### **Federal Energy Regulatory Committee (FERC)**

An independent agency that regulates the interstate transmission of natural gas, oil, and electricity in the United States.

### **North American Electric Reliability Council (NERC)**

NERC members ensure reliability of the integrated North American bulk power system. NERC develops standards, guidelines and criteria for assuring system security and evaluating system adequacy and has largely succeeded in maintaining a high degree of grid reliability throughout North America.

### **Reliability**

The degree of system performance that results in electricity being delivered to customers within accepted standards. Electric system reliability can be addressed by considering three basic and functional aspects of the electric system **Adequacy**, **Security** and the system wide aspects of **Quality**.

- **Adequacy**  
The ability of the electric system to supply electric energy to customers at all times.
- **Security**  
The ability of the electric system to withstand sudden disturbances such as electric short circuits or unanticipated loss of system elements.
- **Quality**  
The extent to which the delivered electricity meets accepted standards such as voltage, frequency and wave shape.

### **System Operator**

The entity responsible for monitoring and controlling an electric power system.

### **Spot Market**

A mechanism whereby electricity generators make competitive offers of specific quantities and prices for electricity at specific times to supply total demand.

### **Automatic Generation Control (AGC)**

**Automatic Gain Control:** A circuit that modulates an amplifier's gain, in response to the relative strength of the input signal, in order to maintain the output power.

**Blackout**

The supply of electricity is discontinued because of outages of generation units or transmission lines.

**Brownout**

A condition where the voltage supplied to the system falls below the specified operating range, but above 0 volts.

**Hourly Ontario Energy Price (HOEP)**

The hourly time weighted electricity price. Ontario market clears every five minutes, and thus there are 12 interval prices with a delivery hour.

**Market Power Mitigation Agreement (MPMA)**

An agreement signed between Ontario Power Generation Inc. and the Ontario government to mitigate OPGI's market power. Under this agreement, the OPGI had to rebate a portion of its revenue to consumers when the market price is higher than \$38/MWh.

**Market Clearing Price (MCP)**

The interval electricity price at which the real time market clears.

**Operating Reserve (OR)**

The combination of spinning and non-spinning reserve required to meet NERC requirements for reliable operation of the grid. These reserves are standing by resources that can be used immediately after a contingency occurs.

**Ancillary Service**

The services other than scheduled energy which are required to maintain system reliability and meet NERC operating criteria. Such services include spinning, non-spinning, replacement reserves, regulation (AGC), voltage control and black start capability.

**Day-Ahead Market**

The forward market for the supply of electrical power at least 24 hours before delivery to Buyers and End-Use Customers.

**Dispatch**

The operating control of an integrated electric system to 1) assign generation of specific generating stations and other sources of supply to effect the most reliable and economical supply as the total of the significant area loads rises or falls; 2) Control operations of high-voltage lines, substation, and equipment; 3) operate the interconnect; and 4) schedule energy transactions with other interconnected electric utilities.

**Forced Outage**

Generating units not in operation due to breakdowns, storms or other unplanned occurrences.

**Locational Market Clearing Price**

The price at which supply equals demand at a specified location. All demand which is prepared to pay at least this price at the specified location has been satisfied. All supply which is prepared to operate at or below this price in the specified location has been purchased.

**Planned Outage**

Generating units not in operation due to planned maintenance, scheduled closures, refueling, or other planned occurrences.

**Ramping**

Changing the loading level of a generator in a constant manner over a fixed time (i.e. ramping up or ramping down). Such changes may be directed by computer or manual control.

**Transmission Congestion**

The condition that exists when market participants seek to dispatch in a pattern which would result in power flows that cannot be physically accommodated by the system. Although the system will not normally be operated in an overloaded condition, it may be described as congested based on requested/desired schedules.

**Wheeling**

The use of the transmission facilities of one system to transmit power and energy by agreement for retail or wholesale customers on another system with a corresponding wheeling charge.

**Baseload**

The minimum amount of electricity delivered or required over a specific time period at a steady rate.