DISPATCH BASED PRICING IN DECENTRALISED POWER SYSTEMS

A THESIS

SUBMITTED FOR THE DEGREE

OF

DOCTORATE OF PHILOSOPHY IN MANAGEMENT SCIENCE

IN THE

UNIVERSITY OF CANTERBURY

BY

B. J. RING

University of Canterbury

1995

Acknowledgments

Contents

List of Figures

List of Tables

Abstract

This thesis investigates the application of marginal cost based spot pricing techniques to the short run coordination of decentralised, and potentially competitive, electricity markets. A Dispatch Based Pricing philosophy is proposed which requires that the dispatcher of a power system determine spot prices which are consistent with both the observed power system dispatch and the offers and bids issued by market participants. Whereas previous research has involved determining prices corresponding to an optimised power system dispatch, Dispatch Based Pricing is more flexible, requiring no such optimality assumption while generating incentives which encourage efficient dispatch. Pricing relationships are formed, and the resulting incentives analysed, by applying duality theory to mathematical programming formulations of the dispatch problem.

A detailed theoretical description of a dispatch based pricing model, based on an Optimal Power Flow formulation, is presented. This model is an extension of the *ex post* pricing model of Hogan (1991). As well as presenting a more general representation of dispatch variable relationships, we demonstrate the underlying mathematical relationships which drive the economic interpretation of this model. In addition, we explore the behaviour of transmission flow constraints in cyclic networks, and describe the modifications needed to price for security requirements consistent with current operational practices in New Zealand.

We explore the extension of Dispatch Based Pricing to situations beyond the scope of the Optimal Power Flow problem, and even to situations which are strictly incompatible with a pure marginal cost based analysis. We develop a "best compromise" pricing approach which, for (seemingly) economically inconsistent dispatches, minimises the side payments required to account for the difference between the market clearing spot prices and the offers and bids of the market participants. We develop and discuss methods for determining dispatch based prices which are consistent with primal inter-temporal constraints, uncertainty, and integer variables.

Chapter 1

Introduction

In recent years there has been growing interest in competitive and deregulated electricity markets. This reflects both a political interest in efficient resource usage and industry frustration, particularly in the United States, caused by imposing overly simplistic regulations on complex power systems. Traditionally, the difficulty of coordinating power systems has limited the ability of power systems to be operated in the decentralised manner required for deregulation and competition. However, Ruff (1992) observes that while large monopolies have traditionally been encouraged so as to take advantage of economies of scale, these economies are actually mainly due to more efficient trade and coordination, and that modern technology can now capture most of these advantages in smaller generating units. This makes it possible, in principle, to achieve market decentralisation, at least in generation, allowing an increase in the scope and efficiency of the market by capitalising on the economies of trade and innovation. This does not mean that no regulatory control should be imposed on the system, but the duties of the regulators should be restricted to those aspects which cannot be efficiently managed purely by the market. In particular, transmission grids are natural monopolies, and should therefore be regulated[1].

Decentralised power systems have no central, all powerful, controller. Rather they involve independent entities all pursuing their own goals[2]. A central coordinator, or dispatcher, may still exist, but is constrained to operate the system in a manner which is consistent with the goals of each entity. Power system decentralisation does not necessarily imply competition or free market access, as the management of the generating

---

[1] This does not mean that the transmission grid plays a secondary role. Indeed, we agree with the opinion expressed by Tabors (1994) that the transmission network, as the provider of the conditions required for competitive trading, should be viewed as the fundamental physical component of an electricity market, rather than the generating stations or loads.

[2] Nor is a decentralised power system necessarily "optimised" by one comprehensive model. Indeed, during New Zealand's shift from a centralised power system to a decentralised system Read (1995) has observed the decline of "all encompassing" market models in favour of less comprehensive models which primarily optimise the factors under the control of the individual market players that operate them.

stations in a monopolistic power system can also be decentralised. However, in general, a competitive electricity market is assumed in this work.

This new structure has created a need for an efficient transmission pricing regime, and particularly one which addresses the externalities which power system transactions create (Hogan, 1992). Due both to the non-linearities inherent in power system operation, and the need to differentiate between physical operation and financial transactions, three basic mechanisms are required for efficient market operations, these being network fixed cost recovery, a short run marginal cost spot pricing regime, and financial hedging contracts between market participants (Hogan, 1992, Read and Ring, 1995d). The short run marginal cost spot pricing regime is related to the problem of the short run coordination of a decentralised power system. This relation arises because the *primal* problem of actually determining how each generating unit or load in the system will operate corresponds to the spot pricing, or *dual*, problem which involves determining the economic value to the system of the net contributions of the generating units and loads. If the dual and primal problems are fully consistent with one another, then either can be used to coordinate the dispatch. It follows that, in a decentralised market, a dispatcher can coordinate generation and load through the use of spot pricing mechanisms, rather than by dictating how each market participant should operate. The fact that demand decreases with increasing price, while generation increases[3], linked with the fundamental power system requirement that total generation matches consumption, implies limits to the range of prices that can be considered by the dispatcher.

The aim of this work is to extend spot pricing theory so as to increase its compatibility with, and applicability to, actual markets for power. This involves increasing the generality and robustness of the theory, while providing a better understanding of the behaviour of prices, and of the forces which drive that behaviour[4].

In many instances, the short run marginal cost pricing theories described in the literature have been designed for usage by utilities which are dominant in their local market. We, however, focus on a pricing systems which can be applied across a national, wholesale market. Past works in this area, particular those of Schweppe et al. (1988) and Hogan (1991, 1992), have approached the pricing problem from a number of different, but

---

[3]   Assuming rational, competitive and unconstrained behaviour.

[4]   While a detailed understanding of power system operation and economics is not assumed in this work, the interested reader might wish to refer to Crew and Kleindorfer (1979), Joskow and Schmalensee (1983), Wood and Wollenberg (1984), Schweppe et al. (1988), US Congress, Office of Technology Assessment (1989), or EPRI (1995).

related, philosophical standpoints.   We explore these approaches and identify some unifying concepts.   We use a mathematical pricing model to highlight a number of relatively simple economic relationships between the power system dispatch and the prices which stem from it.   We also demonstrate that strict adherence to the ideal of "optimality" is not crucial when determining prices.   These ideas should go some way towards increasing the applicability of short run marginal cost pricing techniques to actual electricity markets.

While proposing a philosophical approach to spot pricing, we also demonstrate these ideas in the context of a model developed for the New Zealand electricity system.   This model was developed for Trans Power (NZ) Limited to allow it to more accurately charge for usage of the New Zealand national grid.   The author has played a role in the development of this model and the interpretation of its output (Read and Ring, 1995b, Read and Ring, 1995c, Read, Ring, and Rosevear, 1995).

In Chapter 2 the theory of short run marginal cost based power pricing is reviewed. The market structures which must exist for short run pricing to work efficiently are also discussed. Many of the philosophical ideas raised in Chapter 2 are drawn together with the introduction of Dispatch Based Pricing in Chapter 3.   Dispatch based pricing places the onus on the dispatcher to determine prices which provide a rational economic interpretation of an observed dispatch.   As well as providing a summary of the Dispatch Based Pricing philosophy, Chapter 3, introduces the basic ideas underlying the remaining chapters of this work.

Chapter 4 presents a detailed description of a relatively general single period dispatch based pricing model, its derivation being given in Appendix A. This model is an extension of the *ex post* pricing model of Hogan (1991, 1992), incorporating more general constraint forms and a choice of independent dispatch variables which is more consistent with standard engineering representations.   In Chapter 4 we introduce the concept of determining nodal voltage prices in addition to active and reactive power prices.   The equations describing these prices, and their economic and physical interpretations are explored. This analysis highlights the features of the dispatch which drive the prices. In particular, we demonstrate that dependent pricing variables correspond to independent primal dispatch variables, and vice versa, while constrained primal variables have unconstrained prices, and vice versa.   These results have significant implications to the way the scope of the dispatchers responsibilities should be defined.   The impact of transmission constraints, both on cyclic and acyclic transmission systems, is presented in Chapter 5. We show that the impact of transmission constraints can be significantly more

complex than often assumed, but that this behaviour is governed by fundamental physical relationships.

Power is not the only commodity that might be traded in a power system. Several ancillary services must be supplied if a power system is to operate reliably. An important commodity is the reserve generation and transmission capacity held to ensure the security of supply in the event of equipment failures. In a decentralised market it is important that incentives exist to encourage market participants to contribute to the system's security requirements, and to adequately maintain the reliability of their equipment. A reserve capacity availability pricing model, which ignores the cost of actually calling upon reserve during a contingency, is discussed in Chapter 6. This model is based on the reserve scheduling rules currently employed in New Zealand.

A major assumption of earlier spot pricing proposals (Schweppe et al., 1988, Hogan, 1992) is that the power system is being operated in an economically consistent manner. It is quite possible that this assumption could be violated due to poor management, or appear to be violated as a result of the practical limitations of the pricing model. In these situations the pricing model of Chapter 4 would have no mechanism by which to accurately explain the observed dispatch. In Chapter 7 we propose a modification of that model which allows "best compromise prices" to be found given a (seemingly) sub-optimal power system dispatch. Best compromise pricing minimises the compensation payments which need to be made[5] so as to reconcile any differences between the observed dispatch and the market clearing best compromise prices. We show that best compromise pricing provides a potential mechanism for avoiding costly disputes between the dispatcher and market players while improving incentives for efficient coordination.

Some of limitations of traditional marginal cost pricing theories include the inability to model all inter-temporal linkages between dispatch periods, stochastic effects, and integer dispatch variables. In Chapters 8, 9, and 10 we discuss the ways in which best compromise pricing might be used so as to address these limitations.

Inter-temporal constraints create linkages between prices in different dispatch periods, and these linkages may be too complex to fully model in practice. In Chapter 8 we discuss three approaches which might simplify the modelling of these constraints. The first approach involves no explicit modelling of inter-temporal constraints in the process of dispatching the system, with generating companies being forced to modify their offers so as to account for their inter-temporal constraints. A second approach would have the

---

[5]   An industry levy might provide the revenue required to fund these compensation payments.

generating companies dictating to the dispatcher the levels at which they wish to generate in each period. The final approach is for generating companies to explicitly inform the dispatcher of their inter-temporal constraints, and make it the dispatcher's responsibility to determine generation levels over time. We interpret and compare the implications of each of these approaches in the context of dispatch based pricing. We also show that a second order incentive may exist for generating companies to exaggerate the form of their ramping constraints so as to maximise their profits.

As for inter-temporal constraints, the accuracy with which stochastic effects can be modelled may be limited by the complexity involved. In Chapter 9 we use the stochastic reserve scheduling problem to explore the modelling of stochastic effects. We use an "Optimal Reserve Targeting" model, based on a model proposed by Caramanis et al. (1987), to demonstrate how dispatch based pricing might be applied to this problem. The "Optimal Reserve Targeting" model determines the economically optimal level of reserve to be carried without blackout being called upon. We discuss the role of the dispatcher, the behaviour and interpretation of the dispatch based prices, and illustrate the factors which determine how much reserve capacity a generating unit should provide.

In Chapter 10 we discuss how the commitment costs associated with integer variables might be recovered. Three general approaches are discussed and compared. The first approach involves forcing generating companies to recover their commitment costs via their offers, while the second approach involves extracting rent from the spot market so as to provide extra revenue to cover commitment costs. The third approach involves using some form of best compromise pricing to minimise the compensation which needs to be paid to cover the commitment costs incurred by generating companies, and, in particular, we focus on a best compromise pricing formulation corresponding to the linear relaxation of an integer, and possibly multi-period, dispatch problem.

1.1 Notation and Notational Conventions

The theory of power system pricing combines engineering, economic, and mathematical programming ideas, and has been approached independently from researchers from various combinations of these backgrounds. This has resulted in a plethora of notational conventions. An attempt has been made in this thesis to satisfy the notations of all these fields, though the need to avoid ambiguity, and the large number of symbols required, limits the extent with which this can be achieved. Our notational conventions are summarised in Appendix B.

It is worth noting here, however, that rather than describing prices as being for power usage over some defined period, such as an hour, we assume an arbitrary, unitless, interval. Hence instead of stating that $5/MWh was paid for 3MWh, the total payment being $15, we simply state that $5/MW was paid for 3MW, the total payment being $15.

Following Read and Ring (1995b), the term "node" is used in preference to the term "bus" in all cases except for the swing bus[6]. This convention is used because some buses (eg three winding transformers) may comprise several nodes (Read, Ring, and Rosevear, 1995). We often assume that there is only one generating unit at a node, or, when there are multiple units, that all the units there behave identically.

A physical generating turbine is referred to as a "generating unit"[7]. A set of one or more such units is referred to as a "generating station". The company that operates one or more generating stations is a "generating company". Consumers of power at a node are collectively referred to as a "load", but are individually referred to as a "consumer". The owner of a high voltage transmission grid is referred to as the "grid owner". The entity which coordinates the dispatch of the power system is called the "dispatcher". In a number of situations we assume that the dispatcher and the grid owner are the same entity.

---

[6]   The swing bus is introduced in Chapter 4.

[7]   This should not be confused with a "unit of power", as this is an increment of power.

Chapter 2

Electricity Spot Pricing: Theory and Practice

2.1.   Introduction

Electricity spot pricing is a means by which electricity can be priced so as to reflect the true incremental costs of producing and supplying it to a particular location at a particular instant in time. This form of pricing is highly responsive to changing market conditions, with prices rising as demand increases and falling as it declines, and so is an ideal means for creating rational incentives for participants in a market. In this chapter the role and use of spot pricing theory is described, with particular emphasis on reviewing the strengths and weaknesses of the major theoretical models proposed for the evaluation of spot prices. We also review the institutional structures required if a spot market is to operate efficiently.

In Section 2.2. we describe a number of the electricity markets around the world. These markets have varying degrees of privatisation and deregulation, and demonstrate the role that short run spot pricing can play, as well as the complexities that can be encountered. The early history of spot pricing theory is discussed in Section 2.3. While simplistic models, producing spatially invariant prices, have existed for some time, it has only been in the last 15 years that sophisticated nodal spot pricing models have been developed. The highly influential MIT model, which attempts to determine spatially varying spot prices in real time, is reviewed in Section 2.4. In Section 2.5 we report on the related Hogan model which determines prices *ex post*. A number of other spot pricing models are briefly discussed in Section 2.6.

Spot pricing theory is based on the "first order" economic incentives required for efficient market operation. An electricity market cannot, however, operate efficiently using only spot prices, as many of the players may actually face significant "second order" incentives which distort the market and alter the behaviour of spot prices. In Section 2.7 we describe the origins of these "second order" effects and review the use of long term contracts and capacity rights to reduce their impact.   Possible avenues for research are discussed in Section 2.8.

2.2.   The Context of Spot Pricing

The ultimate aim of spot pricing is to facilitate greater efficiency in the operation of power systems by providing incentives for efficient behaviour by market participants, and hence allowing some degree of decentralised management. It therefore seems appropriate to describe some of the major power markets in which decentralisation is a topical issue. This discussion serves a number of roles. Firstly, it introduces the issues which spot pricing is intended to address. Secondly, it provides insight into the types of market structures which are required for successful spot pricing implementations. Thirdly, it provides evidence that decentralised markets based, at least approximately, on spot pricing philosophies can operate successfully. The theory behind many of the market structures described in this section will subsequently be reviewed.

Of the markets we review[8], the markets of Norway, Chile, and Britain have decentralised operations with a significant degree of competition. The United States market has traditionally been dominated by heavily regulated regional monopolies, but competitive forces are now being seen as a way of reducing regulation and improving efficiency. The Australian market provides an interesting case study as it comprises a number of independent markets which are now becoming inter-connected.   The New Zealand market has been deregulated and is becoming increasingly decentralised. In Section 2.2.7 we summarise the features of these markets which are relevant to the coordination of decentralised markets.

2.2.1. The Norwegian Market

Hjalmarsson (1992) reports that until 1991 the Norwegian power system comprised approximately 600 mostly small generating plants divided among about 63 production and wholesale companies, with about 214 distribution utilities, many of which were vertically integrated. About 99% of the power is produced by hydro generation and, since the country has approximately 7,000MW of excess capacity, it is a major exporter of power. Norway has, for many years, operated an informal, cooperative power pool which uses simple spot prices to assist in coordinating dispatch and market transactions. Holtan, Wangensteen, and Livik (1994) report that reforms since 1991 have encouraged greater competition. The operation of generating stations and the transmission network have largely been divorced and a state owned transmission company, Statnett, has been formed which either owns or leases all parts of the transmission networks. Importing and

---

8   This list is not comprehensive.

exporting monopolies held by Statkraft, the state owned power company comprising 30% of the generation capacity, have been abolished. Local government operates another 55% of the generation capacity. Local government power companies may be vertically integrated, but are required to have separate accounting practices for transmission and generation.

The power pool sets spot prices based on bids from buyers and offers from sellers. The hours of the following day are grouped into similar blocks, there typically being seven blocks, and the buyers and sellers effectively signal supply or demand curves to the pool for each block type. In practice the generating companies must bid separately for each of a number of geographic areas so as to reflect the potential impact of transmission congestion. Transmission congestion can result in the pool being treated as two or more regional pools, each pool having is own local spot price. In addition to the spot market, an access fee is paid by network users so as to cover those networks costs beyond what can be recovered from rents earned from inter-pool transmission.

Holtan et al. report that what are effectively bilateral financial "contracts for differences" between power sellers and buyers are commonly used. These account for about 85% of the power sold, with extra power requirements being traded on a spot market. These bilateral contracts provide a financial hedge against pool price variations (Schweppe et al., 1988, Henney, 1994). For example, a generating company may agree to supply a customer with some quantity of energy at a fixed price at a fixed time. The contract price reflects the players' expectations of what the price will be at the time the trade occurs. At that time both parties trade on the spot market at the prevailing spot price. As the spot price will generally differ from the contract price one party will be better off than expected, while the other will be worse off. The party which is better off then pays its windfall to the other party, with both parties therefore seeing an effective price equal to the contract price.

For instance, if the generating company sells its power on the spot market at a price of β^{spot}, where $\beta^{spot} > \beta^{contract}$, then it makes a windfall of $\beta^{spot} - \beta^{contract}$ on each unit sold. However, this exactly equals the loss made by the consumer, relative to the expected contract price, and by paying this windfall to the consumer both parties see an effective price of $\beta^{contract}$. As observed by Schweppe et al., such long term contracts allow bilateral trades to be defined financially, even though they cannot be defined physically. V. Smith (1993) describes this feature as the ability to define divisible rights for indivisible facilities.

2.2.2. The Chilean Market

Chile was among the first countries to reform its electricity industry to promote greater competition. Bernstein (1988) and Bernstein and Agurto (1993) report that the key reforms occurred in the late 1970's and early 1980's. Spiller and Martorell (1992) report that while there were previously two vertically integrated companies, there are now 11 regional power generating companies, with a cumulative capacity of approximately 18,000GWh, and a regulated, independent, transmission network, comprising 21 distribution companies, not all of which are interconnected. New generating companies have unrestricted access to the market. The market is coordinated by a body known as the Economic Load Dispatch Centre (ELDC). The Comisión Nacional de Energá (CNE) is a regulatory body which develops and coordinates investment plans, regulates prices and acts as an observer and arbitrator for private companies.

Prices at each node of the transmission network are set equal to the local short run marginal cost. The differences between nodal prices is taken by the transmission network operators to cover losses and operational costs. Transmission capital costs are recovered with an annual access toll. Bernstein (1988) reports that the generating companies operate purely on the basis of the spot market.

2.2.3. The British Market

The British electricity market was reformed in March 1990 with the re-structuring and partial sale of the formerly public sector Central Electricity Generating Board (CEGB) (Einhorn, 1995, Green and Newbery, 1992). While nuclear power stations remained under state ownership, generation assets were divided into two geographically dispersed private sector companies, National Power and PowerGen. Entry to the market for generating companies is now open. The transmission network is operated by the National Grid Company (NGC) which is jointly owned by the regional supply companies. The demand side of the market is being deregulated over a number of years, with full deregulation scheduled for 1998. Green (1994) reports that a limited degree of demand side bidding has already been established. A regulatory body, OFFER, acts to protect customers and to promote competition (Littlechild, 1994).

The NGC dispatches the generating units based on price offers for active power issued by the generating companies. The final "spot" prices, calculated half-hourly, include an energy component related to the highest accepted price offer, and a capacity component to encourage generating companies to make their units available at the peak demand times (Tenenbaum et al., 1992). This capacity component is based on the "loss of

load probability" (LOLP), which describes the probability of not satisfying demand, and an estimated "value of lost load " (VOLL). The prices charged to consumers exceed the price paid to generating companies, this difference, or "uplift", paying for many ancillary services such as frequency regulation and reactive power provision, as well as VOLL. At the time of the establishment of the market the uplift was calculated *ex post* so as to recover all the various costs incurred, but Green reports that there is now a move towards it being determined *ex ante* so as to give the NGC a greater incentive to control these costs. All other price components have always been calculated ahead of time, and hence can be inconsistent with the actual dispatch. A fee is charged for access to the network, though this fee does vary according to a measure of network usage.

The rapidly varying spot prices create financial uncertainty for the players in the spot market. In a similar manner to the Norway situation, contractual arrangements between buyers and sellers of power are widely used to hedge against these price variations.

2.2.4. The United States Market

The USA has a very large power system, with over 3,200 electricity suppliers, a total installed capacity exceeding 700,000MW, and annual production in excess of 2,500,000GWh (U.S. Congress, Office of Technological Assessment, 1989). This system comprises many vertically integrated utilities.

Many of the utilities form "pools" which are dispatched as a single entity. The operation of these pools varies. Tenenbaum, Lock and Barker (1992) describe the operation of the New England Power Pool (NEPOOL) which has over 90 members and a total generation capacity of 25,000MW. Pool members are required to install or purchase sufficient generating capacity to meet the peak demand of their retail and firm contractual customers. The pool is centrally dispatched so as to minimise costs. Competition arises through trades in capacity so that the pool members can satisfy the requirement to meet their peak demand. These trades only occur on paper and do not interfere with the physical operation of the pool. Wilkinson (1989) observes that while some degree of competition does result from these transactions, the cooperation involved could be considered to be collusion in a truly competitive market.

The US system is heavily regulated, with both federal and state regulatory bodies. Utility revenues are typically set so as to match operating costs plus a specified rate of return on capital costs. There are also many, so called, Non-Utility Generators (NUGs) from which local utilities are required to buy power. NUG generation is often produced as a by-product of industrial processes.

Regulations are imposed on transmission services so as to encourage efficient usage of the transmission system on a national scale. It has however proved difficult to achieve efficiency in this instance. A major reason for this difficulty is that power flows may traverse the networks of many different, and independently managed transmission companies. These transmission companies will use varying methods to charge for access to, and usage of, their grids. This variety makes the role of the regulators difficult. This situation is further complicated by the so called *loop flow* phenomenon (Hogan, 1992). Loop flows arise because the path that power takes through a transmission network is dictated by the laws of physics rather than by the wishes of the participants in the trade. The existence of loop flow means that any bilateral trade of power between two parties is likely to influence the costs of parties not involved in the transaction. Such externalities make it difficult to efficiently price for such bilateral transactions. Tenenbaum (1993) reports that, in the US, the variable costs of transmission tend to be recovered by prices based on either average or marginal losses, though these calculations are not generally transaction specific, while capital costs are generally based on notional "contract paths".

The difficulties associated with managing inter-utility transmissions and coordinating purchases from NUG generating units primarily arise from the lack of an efficient mechanism to determine fair prices while guaranteeing fair market access. While regulation provides a framework in which these transactions can take place, significant inefficiencies may result. For instance, the California Public Utilities Commission (1995) has announced plans to move towards a deregulated electricity market, a move made primarily as a response to the inefficiencies produced by regulation, which have contributed to electricity prices in California being 50% higher than the US national average. The majority view of the commission favours an approach like that in the UK. A regulated independent system (or grid) operator would control all transmission assets and coordinate a single pool, dispatching generating units according to their owners offers and paying them a uniform market clearing spot price. Customers would have the option to buy power either on the spot market or at a time averaged rate.

2.2.5. The Australian Market

Australia is a large continental land mass with widely spread communities. A single national transmission grid has not yet developed due to the large distances between populations. Instead, each state has owned and operated its own power system, with significant variation in industry structure. Traditionally, there has only been a limited degree of inter-connection between the power systems of the different states, but this is now changing with move towards establishing a national grid. Outhred and Kaye (1994) report that a National Grid Management Council (NGMC) was established in 1991 to

encourage free trade in bulk power and more competitive markets for power. The formation of a grid, with the intent of allowing free trade, has created a need for some framework for trading power between the power systems of the different states.

The NGMC has proposed that a single power pool serves all the inter-connected states, with spot prices being determined each half-hour based on offers submitted by generating companies 24 hours in advance. Market participants could form financial hedging contracts to protect against price volatility.

2.2.6. The New Zealand Market

The New Zealand power system has been undergoing an evolutionary reform process since 1987. A detailed discussion of these reforms is provided by Culy, Read, and Wright (1995). On July 1, 1994 the national transmission grid company, Trans Power New Zealand Limited, was separated from the dominant state generating company, the Electricity Corporation of New Zealand Limited (ECNZ) which produces 96% of the power. Both the generating company and transmission company are operated as State Owned Enterprises (SOE), operating on a commercial footing. Some of ECNZ's generating stations will be split from it to form a new SOE in 1996, with some small generating stations being sold. Entry to the market, both for generating companies and bulk electricity buyers, has not been restricted. Many retail/distribution companies, which were all previously run by municipal authorities, are now publicly listed companies.

At present ECNZ sells its power either on a one year contract or on the spot market. Kennedy (1994) reports that, each year, ECNZ's customers, being local distribution companies or very large industrial users, announce to ECNZ a demand level for each half hour of the following year. For each of these half hours they are allowed to enter into financial contracts for between 90% and 110% of the stated amount at a Time of Use (TOU) rate announced at the beginning of the year. Since 1993 the Electricity Market Company (EMCO), a joint venture between ECNZ and the retail/distribution companies, has facilitated the trade of these contracts, though they may also be traded bilaterally. Every Friday ECNZ issues spot prices of electricity for every half hour of the following week. These prices effectively indicate the expected incremental operating cost of supplying power in each of those half hour periods. Any shortfall or surpluses in the annual contracts can be traded at these prices. An additional "pool price margin" is added to the costs of all units sold to give the final market price. This extra revenue recovers fixed charges and return-on-capital requirements, and is based on a ten-year rolling average of consumption (Culy et al., 1995).

In 1984 a transmission price differential was introduced between the North and South Islands. More recently this concept has been extended to all nodes in the transmission network, the differentials being annual weighted averages of the half hourly spot price transmission components determined by a model similar to that discussed in Chapter 4.

2.2.7. Summary

A common theme in the debate on electricity market decentralisation is summarised in a report by the U.S. Congress, Office of Technology Assessment (1989) in which a need for competitive reform is recognised, but concerns are expressed as to how to efficiently coordinate the bulk power system when many different companies, owning both generating stations and transmission systems, are involved. In this section we review the features of the above case studies which contribute to addressing these concerns.

The operation of what is effectively a single, regulated transmission network (in each isolated power system) is a common feature of the New Zealand, Chilean, British, Norwegian, and Australian models. When generation and transmission are vertically integrated it is theoretically possible to achieve perfect coordination, but competitive pressures for efficiency and innovation suffer. The creation, in each physically isolated region, of a single transmission network under independent control removes the ability of generating companies to control the transmission access and charges of their competitors.

Compared to traditional pricing regimes which do not vary over time or space, the spot pricing approaches used in New Zealand, Chile, Britain, and Norway enhance the ability of independent generating companies and loads to operate in an economically coordinated manner. There do appear to be some limitations inherent in the way in which spot prices have been applied, though. Ruff (1992) identifies two major problems with the way prices are calculated in the British system. Firstly, divorcing the calculation of the pricing components from the actual dispatch has enhanced the ability of some generating companies to manipulate the prices, an ability increased by the lack of competition due to there only being two major generation companies. The latter point is supported by simulations by Green and Newbery (1992). Green (1994) reports that the sale in 1994 of 6000MW of capacity from the two major power companies, and the introduction of 2600MW of capacity from new entrants, all combined with new bidding rules, has gone some way to eliminating this problem, while Newbery (1995) observes that contracts for differences have had a moderating effect on the incentives to manipulate prices. Secondly, Ruff observes that the pricing of ancillary services, which can be vital, is not particularly sophisticated. Green adds that the costs of transmission losses, which he estimates to be

worth £2 million annually, are not modelled, even approximately[9]. Similarly, the spot pricing algorithms used in Chile and Norway, while they attempt to account for the underlying economics, do not accurately account fully for the physics of power systems. These problems expose a limitation in the ability to calculate spot prices rather than any faults of spot pricing itself.

Bilateral contractual arrangements are major features of US power pools and the New Zealand, British and Norwegian markets. These contracts play a major role in transferring risk, which in turn brings stability to market operations. Most transactions take place on the contract market, but the value of these contracts, and the values of deviations from them, are determined by the spot price.

These case studies demonstrate a trend towards separate transmission networks operating as a common carrier for a range of competitive generating companies and loads. The extent to which this market structure can operate without regulation depends to a great degree on the pricing regimes used. Short run spot prices, linked with some form of longer term financial hedging contracts, appear to be the most widely practiced, or at least recommended, regime to clear the energy market, while a fixed network access charge[10] is favoured to recover those costs of operating and maintaining the network which are not covered by rents earned on transmitting power. A clear problem however, and one which this work attempts to address, at least in part, is the problem of how to account for the complexities of power systems within the pricing regime in a manner which is tractable, both in an operational and commercial sense.

2.3. The Origins of Electricity Spot Pricing

Coase (1970) traces the origins of the large body of literature on marginal cost pricing as applied to public utilities back to the 1930's. A major motivation for the development of short run marginal cost (SRMC) pricing techniques was to overcome the inefficiencies of fixed tariffs which fail to provide incentives for efficient energy usage. As marginal costs change with changing power system usage they can reflect the real instantaneous cost of power usage, and hence encourage users to modify their actions accordingly. These can lead to a reduction in peak capacity requirements and higher

---

[9]   Nor do spot prices explicitly account for transmission congestion. Uplift is used to pay generating companies to operate in a manner which enables the system to work around these constraints.

[10]   The UK is unusual in this regard, as the access charge is related to usage.

utilisation of plant. Some utilities, most noticeably Electricité de France (Nelson, 1964), made significant use of marginal cost pricing theory during the following thirty years.

An alternative to SRMC pricing is long run marginal cost (LRMC) pricing. While SRMCs assume all capacity is fixed, LRMCs assume capacity can be varied. Vickrey (1979) opposes the use of LRMC pricing because, for electricity industries, LRMC curves are difficult to define, while SRMC costs can relatively easily be measured and can account for long term construction costs if shortage costs are properly accounted for. In general most power systems are in a state of dis-equilibrium, an effect magnified by the long lead times involved in building new plant. For example, in the late seventies, Mitchell, Manning and Acton (1978) noted that, at that time, most modern power systems were designed in times of low fuel costs, but were being used in a period of high fuel cost. The reverse may well hold for many power systems in the nineties. This dis-equilibrium is a major violation of the assumptions of LRMC pricing theory. Della Valle (1988) suggests that poor investment decision resulting from inaccurate LRMC estimates could be large. She suggests that it is better to use SRMC prices while allowing investors to base their decisions on future SRMC price forecasts. This is in fact the approach used in New Zealand since 1984.

Vickrey (1979) observed that pricing inefficiencies increase as a function of the interval between price changes, and argued that a pareto-optimal[11] pricing system would be one in which the price is updated in real time. Vickrey calls such prices *responsive prices*, though they have subsequently become known as real-time prices or spot prices. The developers of what we refer to as the Massachusetts Institute of Technology (MIT), model, Schweppe, Caramanis, Tabors, and Bohn (1988) recognised the ability of such prices to act as a coordination mechanism in a decentralised market. They argued that, apart from recovering costs, such prices must encourage efficient investment, while also accounting for the constraints imposed by the engineering requirements of power system control, operation and planning. Furthermore, market participants must be free to choose the amount, cost and reliability of the power they use or supply. Finally, there should be no cross subsidisation between market participants. These conditions simultaneously create good incentives for efficient power usage, while encouraging competitive management of power systems. The key issue, though, is how to translate these conditions into a practical model. In the following sections we review models designed to determine spot prices consistent with these conditions.

---

[11]   That is, a price system under which no-one could be better off without making someone else worse off.

2.4. The MIT Model

2.4.1. Origins of the MIT Model

The MIT, or Massachusetts Institute of Technology, model implements many of the ideas brought together in Vickrey's work, and has become the foundation of much of the recent work on electricity spot pricing. The basic model is discussed in this section while some of the more specific details will be discussed in later chapters. A comprehensive description of the MIT model is given by Schweppe et al. (1988).

The MIT model evolved from the concept of *Homeostatic Utility Control*. Schweppe, Tabors, Kirtley, Outhred, Pickel and Cox (1980) describe homeostatic utility control as a price based means for maintaining a state of equilibrium within a power system. A set of SRMC based nodal spot prices, responding in real time to the changing state of the power system, provide incentives for supply to match demand in an efficient manner. These prices reflect the realities of network operation, in which optimal dispatch implies that the marginal cost, and hence the SRMC price, at every node must differ, to a greater or lesser extent, from that at other nodes, in a way which exactly reflects the marginal losses, and any constraints, between nodes. These spot prices would be communicated, in real time, to the loads and generating units which would, ideally, be designed to automatically adjust their operations accordingly. The differences between spot prices at any two nodes would be the transmission price for transactions between these nodes. The transmission prices go to the grid owner to cover the costs of losses and congestion[12].

This approach has the major conceptual advantage that it properly reflects the "pool" like nature of the electricity market, in which transmission costs can not generally be attributed to any specific "transaction". All that can be done is to determine the total cost as a function of the actions of all market participants simultaneously, and define prices reflecting the marginal impact of each participant's actions on that total. Among other things, economic consistency implies that, at those prices, a buyer will be indifferent as to where power is bought from. Buying power at a local node, at the price applying there, will be exactly equivalent to buying power at a distant node, and paying the transmission

---

[12]   As losses are related to line flow by a quadratic relationship, the marginal cost of losses will typically exceed the average cost of losses, hence the grid owner may earn more from the loss component of the price than is required to pay for losses. The congestion component would go towards paying for an increase in line capacity so as to reduce congestion.

charge to deliver it to the local node[13]. Consequently, any number of inter-nodal transactions can be netted out to form a single equivalent node-to-node transaction, independent of any implied "transmission path" through the network. Thus this pricing regime, and only this pricing regime, can create a completely level and friction-less "playing field" for generating companies and consumers to trade and compete on the transmission grid[14].

Caramanis, Bohn and Schweppe (1982) list the following advantages of this approach over more traditional flat rate tariffs:

- Reduced usage of fossil fuels due to more efficient power usage.

- Lower generating unit loadings, hence less peaking plant and lower capital costs.

- Self-rationing of power by consumers during emergencies, and hence a reduced need for direct load control and fewer forced power outages.

- The prices decentralise the dispatch process allowing easier coordination of cogenerators and alternative technologies, and even deregulation.

- The use of "real-time" spot pricing eliminates uncertainty from the price setting process.

- Higher investment in load shifting equipment is encouraged.

- Greater average profits for the generating companies and consumers result.

2.4.2. Spot Pricing with the MIT Model

Caramanis, Bohn, and Schweppe (1982) developed the pricing model which is the heart of the MIT approach. This was the first theoretical mathematical model to describe the detailed form of time varying nodal spot prices in a transmission network. Demand and generating unit availability are assumed to be stochastically distributed with expected prices determined for each time period of the planning horizon. All equations are assumed to be continuous and differentiable.

---

[13]   If this were not so, the corresponding dispatch could not have been optimal. It could have been improved by producing more locally and less at the remote node, or vice versa.

[14]   Ignoring second order effects.

The prices are determined through a price decomposition of the dispatch process. A master problem describes the maximisation of welfare subject to (simplified) power system constraints, while a number of sub-problems determine the profit maximising operating level, subject to constraints, of each player. The optimum spot prices are those prices which are simultaneously consistent with optimal solutions of the master problem and all the sub-problems. It is shown mathematically that for each period and for each node in the transmission network:

Optimum nodal spot price   =   Marginal fuel cost

+   Loss premium

+   Transmission constraint premium

Here the marginal fuel cost, or system lambda, is the fuel cost of the marginal generating unit, i.e. the generating unit which will supply the next unit of power[15]. The loss premium recovers the cost of the marginal losses incurred in sending power from the marginal generating unit to this node. This value is just the marginal transmission losses associated with supplying power to that node multiplied by the system lambda. Finally, the transmission constraint premium alters the price of power when the system line flows are in constraint. These price adjustments may be positive or negative.

Optimal spot prices respond to changing conditions in the network, with Caramanis et al. (1982) concluding that:

> *"Spot pricing does away with the need for any knowledge of individual customers by giving them monetary incentives to act efficiently on their own."*

Rather than having the central dispatcher set usage for each individual user, the users set the dispatch themselves, based on their local spot price, and hence, it is argued, less information exchange is required between the dispatcher and the market about random conditions, nor does the dispatcher need direct access to individual's profit and cost functions. Thus spot prices make it possible, at least in principle, to achieve an optimal dispatch without any centralised physical control. This, of course, would require that spot prices be updated continuously for all customers, and that these customers respond

---

[15]   If there is more than one marginal node then there must be binding constraints on the dispatch. The use of the term "system lambda" can become confusing in this situation, but an analog of the system lambda can be defined as the price at any one marginal node, with the shadow price on the binding constraints explaining the difference between this and the prices at the other marginal nodes.

instantly. The proponents of the MIT model concede that this may not be viable in practice and suggest clustering customers, and using less frequent pricing intervals. Caramanis et al. also conclude that marginal cost based revenues may not match required revenues, and that the steps required to compensate for this will "muddy" the theory.

Caramanis, Roukos and Schweppe (1989) describe the program WRATES, a policy tool for developing rates for the "wheeling" of bulk power between independent utilities. The wheeling rates are based on nodal price differences. This program is based on a simplified representation of a power system and uses different dispatch scenarios to build up a distribution of annual transmission charges.

Caramanis (1982) presents a variation of the basic MIT model which demonstrates the impact of spot pricing on investment. This model differs from the model of Caramanis et al. (1982) by allowing the capacities of generating units to be modified, at some cost, over time. The resulting spot prices have an additional investment term reflecting the value of extra capacity at each node[16]. These additional terms also serve to ration demand in the short term. Hence, if a transmission network has low reliability, then the investment terms will ration demand while providing revenue to improve the network. Incremental investments should be made so as to make the present value of the expected investment related income stream equal to the cost of that investment. It is claimed that the rapid updating of the spot prices eliminates the impact of uncertainty on investment decisions. These last two results assume, unrealistically, that capacity can be expanded or contracted incrementally and instantaneously. Bohn, Golub, Tabors, and Schweppe (1984) explore the conditions required if optimal investment is to occur in a competitive spot market. Free market entry and protection against re-imposed regulation are clearly required if competition is to emerge. They argue that there should be many relatively small generating companies, a conclusion seemingly supported by the case of the UK market[17]. Their final condition, that generation and transmission expansion should be coordinated, is open to interpretation. In principle, this coordination could be achieved either by decentralised market mechanisms or by some centralised authority.

Caramanis, Bohn, and Schweppe (1987) (also see Schweppe et al. (1988)) investigate the integration of system security constraints into their previous work on the

---

[16]   This modification is not required if shortage costs are correctly represented in the short run.

[17]   In essence they argue that no generating company should have market power. The work of Scott and Read (1995) suggests that if most power is sold under contract, with only small amounts of power actually being traded on the spot market, then market power can be greatly reduced, even with a relatively small number of firms.

MIT model. Berger and Schweppe (1989) investigate the generalisation of reserve pricing models to handle time periods measured in seconds. These models are discussed further in Chapter 9. Tabors et al. (1989) discuss early utility experience with marginal cost pricing, while Tabors and Caramanis (1994) review the modelling requirements of SRMC spot prices.

2.4.3. The Practical Limitations of the MIT Model

While the MIT model provides a solid theoretical foundation for spot pricing, it does have practical limitations. Firstly, the technical difficulties of communicating prices and observing the market response to them are great. While not invalidating the concept, these barriers, or more particularly the cost of overcoming them, mean that less ambitious implementations should be considered. Another limitation of this model is that, while it is possible to develop a real time pricing model which describes a power system in great detail, the need for computational tractability are likely to require the use of approximations which will have detrimental effects on the quality of the prices.

Littlechild (1991) investigated the central proposition of the MIT pricing model, namely that spot prices allow the operation of the power system to be decentralised while still producing an identical dispatch to that identified by a single welfare maximising entity. Littlechild concluded that the non-convexities and discontinuities in real welfare functions and power system constraints[18], combined with differing perceptions of the probabilities of stochastic phenomena, mean that prices are not likely to be socially optimal in practice. Schweppe et al. (1988) partially address this issue by stating:

> *"The fact that the true* (prices) *may not be calculated does not destroy the value of implementing a spot price based energy marketplace. The actual value calculated will be much closer to the true values than the present-day flat or time-of-use rates, etc. The goal of implementing the spot price based energy marketplace is to improve the coupling between the utility and its customers, not to achieve theoretical optimality."*

Littlechild also concludes that the multi-period, stochastic spot pricing model described by Caramanis et al. (1982) would suffer from the so called, "curse of dimensionality". That is, the number of states the system can take increase exponentially

---

[18]   For example the energy conservation constraint (see Chapter 4) is a non-linear equality constraint, but to satisfy convexity conditions it must be linear.

with the number of time periods. Furthermore, Outhred et al. (1988) observe that the inter-temporal links in this model are simplistic, and that the model incorrectly assumes that all investment decisions are made at a fixed moment in time, and that there is no uncertainty in construction lead times and project costs.

2.5. The Hogan Model - Ex Post Pricing

2.5.1. Origins of the Hogan Model

Hogan (1991, 1992) proposes a pricing model which can determine nodal spot prices, for a single dispatch period, which can be used for settling market transactions in a wholesale electricity market. A key feature of Hogan's model is that prices are determined after the dispatch has been observed, eliminating the need to solve the primal dispatch problem, hence facilitating detailed representation of very complex systems. This approach has become known as *ex post* pricing.

Hogan originally proposed the use of *ex post* pricing as a means of resolving many of the commercial difficulties arising from long distance inter-utility transmission in the United States[19]. These difficulties arise because traditional pricing practices are based on ad hoc rules (e.g. contract path or megawatt-mile) which fail to account realistically for the complex physical behaviour of power systems, particularly loop flow[20], and which therefore create inadequate incentives for the players. Hogan argues that an efficient transmission market should be based on long term transmission capacity rights with short run users being forced to recognise the opportunity costs implied by their actions. These opportunity costs would be communicated via prices which would, therefore, promote efficient trade over short time intervals.

While Hogan endorses the general approach of the MIT model, he rejects the homoeostatic control concept upon which that model was based. Hogan concludes that, given the power system control mechanisms already in place, homeostatic control is unnecessary and that it may undermine the reliability of a power system.

---

[19]   Stalon (1987) discusses the problems that these difficulties create for regulators.

[20]   "Loop flow" describes the fact that power flows between two points do not follow the shortest path, or a contractually agreed path, but may be split among any number of different paths in a way which minimises the total impedance experienced between the two points. Baldick and Kahn (1995) raise the possibility of using phase-shifting transformers to allow some degree of control of power flows, but we do not consider this possibility.

By determining prices given an observed dispatch Hogan's approach can model the full complexity of an AC representation of a power system. While the MIT model could be applied to such systems, the need to simultaneously solve the primal and dual problems may limit the complexity that can be represented in practice. Hogan argues that it is important to model a full AC system, as it can be shown that reactive power prices must be represented if active power prices are to fully reflect economic costs (Hogan, 1993, Baughman and Siddiqi, 1991)[21]. A further feature of *ex post* pricing is that it actually reduces the risk of price manipulation in competitive markets. If nodal prices calculated before the event (*ex ante*), are applied to actual (*ex post*) volumes there may be some potential for generating companies to manipulate the price by issuing misleading information about their availability[22]. *Ex post* prices go some way towards eliminating this form of price manipulation as they are based on what actually happened, rather than what was expected.

2.5.2. Spot Pricing with the Hogan Model

Hogan presents a model for determining the nodal active and reactive power prices in an AC power system. Hogan's pricing equations are similar to those of the MIT model but are derived from the dual of an AC Optimal Power Flow (OPF) formulation. Hogan assumes a welfare maximising objective function. The dual OPF is formed from a linearisation about an observed power system dispatch. The prices can be found by solving this dual formulation after substituting into it the marginal costs, the incremental impacts of players on each constraint, and the binding constraints corresponding to an observed dispatch. The fundamental assumption of this approach is that a power system is operated in an economically efficient manner and that economically consistent prices can therefore be deduced.

The *ex post* pricing regime can be generalised to assume that the dispatcher "optimises" the system using the supply and demand curves implied by pre-dispatch offers from generating companies and bids from wholesale customers. In practice this is likely to involve a "pre-dispatch" process to determine planned, and expected, operation perhaps a day ahead (as in the UK) or a week ahead (as in Norway), plus real-time control to adapt that plan to meet changing circumstances. Ruff (1992) observes that, as long as the

---

[21]   Kahn and Baldick (1994) note, though, that the magnitude and impact of reactive power prices may be exaggerated if the observed reactive power generation and line flows have not been optimised.

[22]   As encountered in the UK system.

dispatch is consistent with these bids and offers, and the prices are consistent with the dispatch, the market participants should be satisfied.

Hogan's AC pricing method has successfully been applied to realistic networks. Rosevear and Ring (1992) report that prices for all 600 nodes in the New Zealand system can be determined within about 90 seconds using a 386 PC. These prices are currently only used for determining fixed average annual nodal price differentials. In the United States, this model has be successfully applied to problems with over 3768 nodes (Hogan et al., 1995)

2.5.3. The Practical Limitations of the Hogan Model

A central assumption of Hogan's approach is that the dispatcher can determine some form of dispatch, in an unspecified manner, from which "optimal" prices can be derived. Given the fact that duality theory imposes relatively strict requirements on the range of feasible pricing solutions that can exist, it may occur that there are no prices consistent with an observed dispatch. Outhred (1994) is particularly critical of this aspect. This issue is explored more fully in the following chapter.

Hogan does not provide detailed analysis of inter-temporal and stochastic aspects, nor does he use a general set of independent dispatch variables. Accounting for these factors has a significant impact on the both economic interpretation and the implementation of such a pricing model. As will be argued in the next chapter, factors such as stochastic effects can require more subtle distinctions than the simple *ex post*/*ex ante* characterisation made here.

2.6. Other Spot Pricing Models

While we primarily use the models of Hogan and Schweppe et al. as our reference models, there is an ever increasing range of publications on related models or applications and extensions of these models to specific problems or situations. In this section we briefly summarise some of these works.

Rivier and Pérez-Arriga (1991) present examples of MIT model spot pricing calculation using their JUANAC model. Baughman, Siddiqi and Zarnikau (1992) discuss an AC extension of the MIT approach. While they provide general representations of a wide variety of constraints, including security constraints, pollution emission limits, and harmonic distortion, they do no explore detail implications or precise form of each constraint, nor do they discuss how such a model should be solved.

Baughman and Siddiqi (1991) use prices derived from an OPF (which includes the modelling of demand elasticity) for a simple 4 bus system to highlight the importance of reactive power pricing. Chattopadhyay, Bhattacharya, and Parikh (1995) use a reactive power pricing model as part of a cost-benefit analysis to determine the size and placement of capacitors in a transmission network. Dandachi et al. (1995) study the feasibility of using a security constrained optimal power flow problem to price for reactive power in the UK, and while noting that further work is required, draw favourable conclusions.

While a detailed discussion of retail markets is outside the scope of this work, the connection between wholesale spot pricing and retail pricing is important, as the full advantages of responsive wholesale pricing cannot be gained unless demand reacts to price at all levels. Chao and Wilson (1987), and Chao et al. (1986) propose the use of "priority service contracts" which allow consumers to select a desired level of service reliability from a menu. Associated with this reliability is a priority charge per unit contracted, and a service charge per unit actually delivered. A utility selling such contracts then has the right to curtail load to customers in ways which are consistent with the service reliability specified by each customer. The priority service contracts held by a customer effectively act as an "expert system" representation of that customers' response to spot prices. However, in determining who is served, it is not necessary to know the spot prices, customers are simply served in order of their specified reliability levels. Chao and Wilson conclude that in the absence of transaction costs, priority service contracts and spot pricing will be equally efficient methods and, given perfect information, will both achieve a socially optimal allocation of resources.

In the context of the traditional market structure in the United States, Siddiqi and Baughman (1992) propose Reliability Differentiated Real-time Pricing, an approach related to priority service contracts, which allows a utility to ration supply by charging consumers different prices, these prices rising with the firmness (or reliability) of supply required, while Siddiqi and Baughman (1994) discuss the application of this approach to the pricing of Non-Utility Generation. For a similar context, Murphy, Kaye, and Wu (1994) derived simplified spot pricing relationships for radial local distribution networks (sub-station level and below).

2.7.   "Second Order" and Long Run Issues[23]

Marginal cost pricing only provides "first order" incentives to players in a competitive market, and cannot be guaranteed to provide correct signals in the long run. In this section we briefly review the mechanisms which can be used with SRMC pricing so as to address these issues.

2.7.1. Network Cost Recovery

Read and Sell (1989) observe that most regions in New Zealand are served by only one or two lines and that the large "fixed" investment costs for these lines imply optimal line capacities well in excess of technical requirements. That is, once the decision has been made to serve a region with a line, and the fixed costs associated with a small line have been committed, the incremental costs of actually installing a larger line are relatively small, and easily justified in terms of the long run impact on losses, congestion etc. Since the expected revenue from capacity rentals, or the "quality of supply premiums" referred to by Schweppe et al. (1988) will, on average in the long run, only equal these incremental costs, a significant proportion of the fixed investment costs will not be recovered in this way. Read and Sell used a DC approximation to estimate that, in New Zealand, only 10% to 30% of total costs of optimally sized transmission lines could be recovered through optimal short run pricing.

More recent experiments with the AC pricing model of the New Zealand system suggest that these DC approximations significantly under-state the cost recovery problem. In fact, the charging currents required by lightly loaded lines reduce the rent which might otherwise be expected from marginal cost pricing for losses, to the extent that the transmission system sometimes operates at a loss, and barely breaks even in many periods. Read (1989) proposes that such a large discrepancy can only be recovered through some type of "fixed charge", "taxation", "club membership fee", or "take-or-pay contract" determined in a way which provides efficient long-run incentives while not distorting the short run incentives. Ideally this latter requirement means that long run charges should be independent of actual usage during the period of the contract, or pricing regime.

Read and Sell also argue that, ideally, transmission facilities should only be built when users are prepared to pay for them because they judge that the net present value of the cost of facing the extra spot price differentials which would eventuate without

---

[23]   This section is based on Read and Ring (1995d) and Ring and Read (1995).

expansion exceeds the capital cost of the expansion. Users should then commit to a take-or-pay contract at the time of expansion, and receive "capacity rights", as discussed below, for the new capacity.

2.7.2. Market Power, Contracts, and Capacity Rights

The SRMC pricing mechanisms described above assumed that all generating companies are price takers. In practice the owners of large generating stations may have significant market power to influence prices. For example, a large low cost generating station could operate at below its full capacity, forcing a more expensive generating units that otherwise would not be operated to cover the difference. The market price must rise to cover the marginal cost of the more expensive units. The extra revenue produced by the higher price being applied to the "gaming" company's output may offset the revenue lost in restricting its capacity, giving it a significant second-order incentive. The theoretical work of Scott and Read (1992, 1995) suggests that these incentives can be greatly reduced if generating companies have a significant proportion of their capacity under long term contracts for differences, a result consistent with the observations of the UK market reported by Newbery (1995)[24].

Most analyses of gaming and contracts have concentrated on "energy" markets, but the way in which losses depend on the actions of other parties, and the potential for transmission networks to become congested, creates significant uncertainty about short run prices at any particular node in the network, and between any pair of nodes. In particular, transmission congestion may effectively segment the market, increasing the market power of some generating stations by restricting market access of others. Heinz (1995) identifies this as a major problem in Argentina, where generating companies must contribute to the costs of operating the transmission network but are not guaranteed access to the network and receive no compensation if they are forced out of the dispatch by congestion. Read and Sell (1989) and Hogan (1991, 1992) have proposed capacity right mechanisms which help to overcome this problem. These rights would be financial contracts, based on the

---

[24]   To deal with problem of profit maximising utility restricting supply to force up prices Vickrey (1979) proposed an *escrow fund*. Consumers would pay a time varying spot price while generating companies would receive a pre-arranged, time-of-day, *standard price*. Any positive (negative) differences between the spot price and the standard price would paid into (taken from) the escrow fund. Thus any generating company that restricts supply to force up the spot price will lose the additional profits to the escrow fund. This approach is too simplistic as the utilities lose many of the operational incentives associated with true spot pricing.

differences in nodal spot prices[25], thus providing network users with a result equivalent to being able to send an agreed amount of power between nodes, on a hypothetical "dedicated line", at an agreed cost. When flows between the nodes defined in a capacity right are constrained, spot prices will move up at the "receiving" end and down at the "sending" end to reflect the capacity charge component of the price. Prices may even fall so low at the generator node that it is forced out of the dispatch, but the capacity right will still refund the capacity component on the agreed quantity, irrespective of the actual generation level. This effectively gives the generating station access to the price at the "receiving" end for that quantity of power generated by it or, if it is forced out of the dispatch, "bought in" from cheaper sources. The rights proposed by Read and Sell go further by refunding the rental component implicit in marginal losses, thus effectively giving access to a notional dedicated line, at the corresponding loss factor. Recently, Chao and Peck (1995) have proposed the use of a DC load flow model to evaluate capacity rights similar to those proposed by Read and Sell, while Bushnell and Stoft (1995) have discussed the allocation and trading of Hogan style capacity rights.

If network users do not hold capacity rights roughly in proportion to their "usage", that is for the amount they wish to buy/sell on the "other side" of the constraint, they may have incentives to distort the inter-nodal differentials. For example, if two users are equally dependent on a line, but one holds all the rights, the one holding the rights may wish to force the line into constraint, by adjusting generation or load at certain times, thus forcing the other party to pay rent. Such behaviour may seem undesirable, but it can also be seen as part of the normal commercial process which gives both parties incentives to pay for, and obtain rights to their "fair share" of the capacity. It can be shown, for rights of the form proposed by Read and Sell, that second order incentives will tend to move actual utilisation toward the capacity right holdings, and so will tend to stabilise usage patterns around these contractual agreements, without imposing any arbitrary or rigid restrictions (Read, 1991).

More subtle effects arise with losses, due to the non-linear (and near quadratic) nature of the loss function. Using a DC approximation, it is easy to see that if a generating company were responsible for the actual losses on its own spur line, it would differentiate the loss function to determine a marginal loss equal to twice the average loss, which is the

---

[25]   Capacity rights differ from "contracts for differences" in that the former are based on differences between spot prices at two nodes while the latter are based on the difference between either a spot price at a single node, or the pool price, and a previously negotiated contract price. Contracts for differences normally only hedge against price uncertainty while capacity rights hedge against uncertain market access.

correct "first order" price differential, such as would be determined from a DC pricing model. However, if it were faced with the "optimal" prices based on <u>marginal</u> costs, it will still face a quadratic cost function (price multiplied by losses) for losses on the spur line, but one which rises at twice the true rate. Differentiation will then yield an effective price differential twice as great as it should be, giving the generating company artificial second order incentives to reduce generation in order to get a higher local price. Read and Ring (1995d) observe that the impact of such effects is debatable. In dense systems, "sole use" of lines will be unusual, and loss differentials will also be small, so that doubling them will make little difference. In thermal systems, distortions will have no impact unless they are larger than the price steps in the supply curve. In hydro systems, prices are determined from water values representing the value of storing water for future production at similarly distorted prices. Thus, ignoring the possibility that spill will increase, there can be no overall impact on production, since water must eventually be used.

To the extent that there is a problem, it can not be overcome by charging average loss costs, as this will only give correct price signals for monopoly users of lines, whereas most users would be minority users of most lines. However, if users hold capacity rights of the form described by Read and Sell (1989), in which losses are accounted for, the second order incentive is modified to move generation towards the capacity right allocation. This generally reduces distortion, removing it entirely if capacity right allocations match optimal generation patterns.

Rosevear, Read and Watson (1992) identify a similar second order problem with loss differentials for reactive power, since optimal reactive power prices at generator nodes will typically be zero. They suggest that in this case the best solution is probably for the dispatcher to buy the right to dispatch reactive generation, within limits, rather than paying for the reactive power on an half-hourly basis. If the transmission grid is managed by the dispatcher, as is the case in New Zealand, then second order incentives can potentially exist for the dispatcher as well, since declaring a transmission line to be constrained increases the dispatcher's revenue. If this is believed to be a problem, though, users should buy capacity rights for the system capacity, since this not only protects them from the effects of such manipulation, but removes any incentives for the dispatcher to act in this way.

2.8.   Avenues for Research

It is apparent from the international case studies that the basic concepts of spot pricing have gained wide acceptance, however there is still a need to refine both the

philosophy and theory of spot pricing. Some potential refinements, specifically relating to spot pricing, which might be made include:

1. Generalising the philosophy of spot pricing to encompass the approaches of Hogan and Schweppe et al., while eliminating the need to assume that the dispatch is actually optimal, as this is unlikely to hold in reality.

2. Increase the generality of the pricing models to include a wider range of dispatch constraints.

3. Increase the robustness of the mathematical pricing model to make it consistent with the generalised philosophy, allowing it, for example, to deal with primal sub-optimality.

4. Devise mechanisms for addressing those dispatch constraints which are either inconsistent with the assumptions of marginal cost pricing (e.g. integer variables) or involve excessive modelling complexity to fully represent (e.g. inter-temporal and stochastic effects).

5. Explore and explain the behaviour of prices in practice.

6. Examine the "second order" incentives implied by these "first order" prices, and, where necessary, develop mechanisms to modify these incentives.

In this work we primarily address the first four points, with a more general "Dispatch Based Pricing" philosophy being proposed in the next chapter, and a generalised mathematical pricing model discussed in Chapters 4 to 6. Chapter 7 deals with increasing the robustness of this model to deal with apparent primal sub-optimality while Chapters 8, 9, and 10 suggest possible ways of addressing inter-temporal, stochastic, and integer effects within the dispatch based pricing framework. While the final two points are not primary goals of this work, we do report on some aspects of observed pricing behaviour to illustrate the importance of some theoretical features, and discuss the second order incentives of some of the new pricing model features we propose.

Chapter 3

Dispatch Based Pricing[26]

3.1. Introduction

In the final section of the previous chapter it was observed that there is a need to develop a broader philosophical view of spot pricing. In this chapter we introduce a new philosophical interpretation of spot pricing called "Dispatch Based Pricing". This philosophy provides a different perspective on the pricing models proposed by Schweppe et al. and by Hogan, with prices being viewed as a mechanism by which the system dispatcher can "explain" an observed dispatch to the market, rather than just reflecting the value of the contribution of each market participant. In many instances the results will be equivalent, but we argue that the dispatch based pricing philosophy provides a more flexible, and hence practical framework within which to explore and interpret spot pricing formulations. We consider the rationale for this philosophy, and discuss some of the practical implications. Further implications will be explored, in the light of mathematical models, in later chapters.

3.2. Dispatch Based Pricing Philosophy

The approaches of the MIT model and the Hogan model have variously been described as "Short Run Marginal Cost Pricing", "Nodal Spot Pricing", and "Ex Post Pricing". Each of these terms describes one aspect of the fundamental approach, but none is wholly satisfactory. A pricing model may include constraints, and produce prices, which are not specifically "nodal", while short run marginal cost pricing could be *ex post* or *ex ante*. Even the precise definition of *ex post* becomes contentious when dealing with dispatches made, perhaps sub-optimally, under uncertainty. Strict *ex post* pricing would be based on dispatch <u>outcomes</u> but, when operating under uncertainty, real time dispatch <u>intentions</u> may be more relevant. The use of the term "marginal cost" may also become contentious when considering non-convexities in cost functions and dispatch constraints.

---

26   This chapter is based on Read and Ring (1994a).

Thus we prefer the more general term "Dispatch Based Pricing", which reflects the basic goal, which is that:

> *Prices should be determined, after the fact, so as to be consistent with the observed dispatch, in the sense that they provide a rational economic explanation for it, given the price "floors" and "ceilings" expressed by the sellers and buyers in their pre-dispatch offers and bids.*

This approach allows all market participants to indicate to the dispatcher what services they will supply or require and the minimum price at which they will supply them, or the maximum price at which they will purchase them. The dispatcher is free to dispatch the system, given this information, in any manner, but must be able to provide an explanation of that dispatch, via prices. In principle, this definition allows us to relax the requirement, implicit in the dual formulations of Schweppe et al. and Hogan, for example, that the dispatch be optimal, with hindsight. Thus a "rational economic explanation" of a dispatch may be that the dispatcher understood, rightly or wrongly, that a particular constraint on the system was limiting the dispatch, or even that the dispatcher was not free, or not able, to optimise the dispatch.

The fundamental requirement for a rational economic explanation is still provided, though, by applying classical linear programming duality theory to the (possibly hypothetical) optimisation performed by a cost minimising dispatcher. Any departure from strict merit order operation can be "explained" by claiming that it was necessary in order to work around some constraint. This "explanation" is provided by the dual shadow price assigned to the constraint, and the fact that all other aspects of the dispatch must be consistent with that shadow price. One of the consistency criteria which duality theory imposes on the prices is that the price assigned to any variable which has been optimised by the dispatcher, and is not forced against a constraint in the observed dispatch, lie between the net cost of decreasing it and that of increasing it. For example, active power prices, at each node, must lie between the bid/offer floors and ceilings for that node. This requirement can ensure that all parties, except perhaps the dispatcher, will be adequately rewarded for their actions. On the other hand, this approach also allows the flexibility to relax the pricing consistency requirements, with the consent of those affected, by assuming that some variables have not been optimised by the dispatcher, or at least not within the time frame under consideration. For example, in New Zealand, the reactive power and voltage magnitude settings for generating units are currently treated in this way, since they

are not (yet) actively optimised within each half hour scheduling period[27]. One critical issue, then, is to determine the scope of the dispatcher's responsibilities, in terms of[28]:

- the variables and resources which the dispatcher is assumed to control and/or "own",

- the variables and resources for which the dispatcher is responsible to optimise utilisation,

- the time frame within which, and over which, the dispatcher is assumed to optimise, and

- the degree of foresight which the dispatcher is expected to exercise.

Once these issues have been determined, the appropriate form of the pricing model follows, as demonstrated in Read and Ring (1995b). Typically, the formulations which have appeared to date have taken a particular view on these matters, requiring, for example, that we assume each observed dispatch to be optimal in terms of the constraints which are evident within a particular half hour snapshot, since otherwise no consistent set of prices can be found. It is obvious, though, that a particular half-hourly dispatch which appears to be sub-optimal, when viewed in such terms, may actually be optimal when seen in a wider context involving ramping constraints and integer unit commitment costs. While practitioners using snapshot models can find ways to account for such effects, the apparent restrictiveness of some of the assumptions has lead to criticism of this type of model. We believe that much of this criticism is mis-directed, and can be avoided if these formulations are seen in the context of the broader dispatch based pricing approach advocated here.

Similarly, criticism has been levelled at the whole concept of *ex post* pricing on the ground that market participants need price signals to guide their behaviour before the event, not after. By advocating that prices be evaluated *ex post,* though, we are not suggesting that no price information would, or should, be available *ex ante.* Rather we expect that, as the dispatch time approaches, pre-schedules will be formed, prices will be estimated on that basis, and deals will be done at those prices, but guided by the knowledge that, finally, the *ex post* prices, reflecting the economic impact of actions by all parties in real time, will apply to deviations from whatever prior deals may have been done. This approach protects the commercial interests of all parties, while allowing

---

[27]   This example is discussed further in Chapter 4.

[28]   These issues are discussed in a mathematical context in Chapter 4.

dispatchers, rather than being constrained by arbitrary "physical" contracting arrangements, to control the system in a secure and economically rational way, knowing that the economic implications of that dispatch will be consistently reflected in prices which deliver appropriate commercial and economic signals to all parties. In this way dispatch based pricing can allow the physical and economic complexity of network operation to be reflected in prices to the extent that the most sophisticated market participants deem to be commercially worthwhile, while allowing, and giving appropriate incentives to, all parties, including the dispatcher, to act in response to the best, and most up-to-date set of prices they find it worthwhile to obtain.

3.3. Explaining an Observed Dispatch

In this section we present examples of how the dispatch based pricing philosophy can be used to provide interpretations of the key features of an observed dispatch. Although we will later characterise the pricing problem as the dual linear programming problem corresponding to the linearisation of the original primal dispatch optimisation, the dispatch based pricing methodology does not require that a pricing model, or the corresponding primal formulation, be explicitly solved during the dispatch process. It is claimed that a dispatch based pricing model may be applied to any observed dispatch, derived in any fashion. This claim is advanced, though, on the assumption that any dispatcher will have found an economically rational dispatch, that is, one which is optimal for the (primal and dual) constraints reported to the model[29]. Thus, rather than concentrating on the *ex ante* "primal" problem which may, or may not, have been solved by the dispatcher, it is more helpful to concentrate on the *ex post* "dual" problem faced by a dispatcher trying to justify that dispatch.

Generally, if the dispatch is in strict merit order, there will be no need for a sophisticated model to "explain" the dispatch. It is only where "out of merit order dispatch" (OOMOD) occurs that such "explanations" are necessary. Thus the whole pricing process described here is driven primarily by the observation of OOMOD, and the attempt to explain it, rather than by direct consideration of the dispatch process and the

---

[29]   In Chapter 7 we will show that this assumption can also be dropped if the dispatcher is prepared to pay compensation for incorrect dispatching.

constraints which may have operated during it. In fact if a physical constraint causes no observable OOMOD, it is irrelevant for marginal cost pricing purposes[30].

In order to focus our discussion it will be helpful to relate the issues to the following simple example of apparent OOMOD:

> The generating unit at node *i*, having a capacity of 100MW and a "marginal" cost of 2c/kW, was observed to run at only 95MW, on average, over a dispatch period. The system marginal generation cost was 2.2c/kW, suggesting that it should have produced 100MW.

As will be seen from the discussion below there are several possible "explanations" of this event. Broadly speaking, these explanations fall into seven categories:

- the impact of losses.

- misrepresentation of the data describing the observed dispatch.

- system constraints.

- inter-temporal constraints.

- changed, or unforseen, circumstances.

- integer effects.

- truly sub-optimal dispatch.

As will be seen, these explanations have significantly different pricing implications, as expressed by the price "floors" and "ceilings" which can be inferred from the observation. We might also observe that the plausibility of the possible explanations varies widely. Fortunately, the more extreme pricing implications are generally associated with the least plausible explanations.

3.3.1. The Impact of Losses

As observed by Schweppe et al. and by Hogan, the impact of transmission losses, both for active and reactive power, can be represented by the marginal change in system

---

30   The converse statement is not true, though, as a constraint which affects the dispatch may or may not effect marginal cost based prices. For example, a non-convex constraint may have an impact on the dispatch, but will not necessarily have any impact on marginal costs as these assume convexity.

losses with respect to a marginal change in net injection at each node. One node, introduced in the next chapter as the *swing bus*, is treated as a nominal source or sink of the power change required to balance the change in losses. At this node the loss derivatives are zero. The price at each node must be adjusted to accounted for the cost of covering the change in marginal losses. Hence the first possible explanation for the apparent example of OOMOD at node i is simply that the loss adjusted price at the node was exactly 2c/kW, and that the unit was deliberately and correctly scheduled between its bounds. The difference of 0.2c/kW between this price and the system lambda represents the cost saving made by reducing losses through generating power at node i, rather than at the marginal node. In this case node i was a marginal supplier of energy[31], implying that the price at that node must exactly equal the unit's marginal cost of 2c/kW. That is we have:

Nodal energy price floor = ceiling =2c/kW

Attempting to explain prices in terms of the cost of power at the marginal source and loss derivatives produces a "loss differential solution". If these prices are consistent with all offers and bids, the pricing problem is solved, but each remaining inconsistency must be due to one of the other causes discussed below.

3.3.2. Misrepresentation of the Data Describing the Observed Dispatch.

Before looking for any other explanation, we should consider the possibility of data error. In fact, minor discrepancies between the primal and/or dual constraints assumed and reality are likely to be quite common, for a variety of reasons. The problems related to node i could include the following:

- The unit generation may have been wrongly recorded, and it actually did generate 100MW. Alternatively it generated as much as it could, which turned out to be 95MW in that dispatch period, due to problems with the unit, or to the dispatch situation changing within the period. Thus the price there could be anything greater than 2c/kW, i.e.:

Nodal energy price floor =2c/kW, no ceiling

---

[31]   Under such circumstances there can be multiple marginal nodes, all with equivalent loss adjusted prices. That is, if prices are set at one marginal node then adjusting that price for losses will produce the correct price at each of the other marginal nodes. Hence the choice as to which node is selected to be "the" marginal node, and hence set the system price, is arbitrary.

- Logically, we should also consider the reverse alternative, although it seems less plausible in this case, that the unit was actually not generating at all, or was generating as little as it could, because of some minimum running constraint[32], in which case:

Nodal energy price ceiling =2c/kW, no floor[33]

- The marginal cost estimate of 2c/kW may not have properly accounted for efficiency effects. Although generation up to 95MW may only cost $c^-=$2c/kW, generation beyond that point might cost, say, $c^+=$2.3c/kW[34]. In that case the unit would have been correctly scheduled at 95MW, and the price at the node can lie anywhere between these two limits, i.e.:

Nodal energy price floor =2c/kW, ceiling =2.3c/kW

Obviously, the extent of any apparent physical "discrepancy" is a critical issue here. The plausibility of the various explanations advanced above varies dramatically as the observed solution varies between, say, 50MW and 150MW.

Automatic checks must obviously be made for data integrity, but dispatcher input will surely be required to determine whether a unit will be deemed to have been fully loaded on any particular occasion. In view of the pricing implications, generating companies and consumers will certainly have conflicting views on these matters, too. Thus it is suggested that the intentions and judgements of the dispatcher should be recorded as close to the dispatch time as possible, and preferably checked for consistency at that time. One way to check the consistency of the intended dispatch would be by forming its corresponding pricing solution.

We also note the necessity to carefully and consistently define exactly how the pricing data is to be collected, whether in the form of snapshots or of integrated measurements over the period. For example, suppose an expensive unit generates at some

---

[32]   Such minimum running restrictions can arise from river flow constraints on hydro generators, for example, or from integer loading restrictions. The latter can be accounted for by setting an appropriate price ceiling in the model, as has been done here, but this does not guarantee a fully consistent solution over time. This issue is addressed under the heading of integer effects below.

[33]   We can not even infer a floor of zero. There are circumstances when nodal prices can become negative.

[34]   In general, marginal cost will rise with output, even if average cost is falling.

time during the half hour. It may be taken to be marginal over the whole half hour if pricing is based on integrated measurements, and this may establish an unrealistically high price. On the other hand a snapshot may exclude the unit all together, forcing it to generate at a loss.

3.3.3. System Constraints

Assuming that the apparent dispatch discrepancy is not explained by loss differentials, or by data error, the most likely explanation will be in terms of some system constraint having prevented a more economic dispatch. While the form and nature of these constraints will be explored in later chapters, all that is relevant to this discussion is that these constraints can be broadly grouped as "electrical" constraints and "operational" constraints.

For purely "electrical" constraints in which the generation and demand variables appear symmetrically (including line and voltage constraints), we have the following explanation:

- That the unit at node i was deliberately and correctly dispatched at 95MW in order to help meet some electrical constraint which was believed to be binding, implying that the price for energy at that node should be set to the marginal cost of 2c/kW, but that the relevant constraint should be invoked to explain any discrepancy between this and the prices implied by other marginal units. Thus:

> *Nodal energy price floor = ceiling =2c/kW*

> *Constraint multiplier activated*

For "operational" constraints, such as emergency reserve capacity constraints, in which power generation plays a fundamentally different role from "negative load", we have a slightly different explanation:

- That the unit at node i was correctly and deliberately dispatched at 95MW in order to help meet some other constraint, implying that the <u>combined</u> price for energy and, say, reserve, must equal the marginal cost of 2c/kW[35].

> *Combined nodal energy price/reserve price floor = ceiling =2c/kW*

---

[35]   Refer to Chapter 6.

Reserve constraint multiplier activated

If the system is optimally dispatched, system constraints should explain most situations not already explained by the loss differentials. Deciding whether a constraint is binding, or not, may not always be a simple matter, though. Transmission constraints, for example, may often be considered as "soft", with a certain amount of "overload" being acceptable for a short period on a cold day, but not in other circumstances. Contingency constraints, such as reserve constraints, are even harder to quantify. Again dispatcher input is essential, and judgements can be expected to be contested.

Obviously, the extent of any apparent physical "discrepancy" is a critical issue here too, as is the extent to which the constraint concerned affected the dispatch. If there is OOMOD, some constraint must be invoked to explain it, and it will have to be assigned a shadow price large enough to create a price differential at which that OOMOD becomes economic. Conversely, if there is no detectable OOMOD, no price will be assigned to the constraint even though the dispatcher may have believed it to be binding. It may be thought that one way of resolving some ambiguities is to specify several constraints as binding, and let the pricing model decide which it will assign a price to[36], but this may not be desirable if the pricing model objective function does not match the true dual objective function and if the dispatch is not actually optimal.

3.3.4. Inter-Temporal Constraints

Hogan's model, unlike some theoretical forms of the MIT model, is implicitly based on a single period, and does not explicitly account for inter-temporal constraints such as ramping requirements on thermal generating units and carryover of hydro storage.

In the latter case, we might expect that hydro system managers will have specified half-hourly water values to the dispatchers, for each station[37], and that these will be used to form price constraints in the model. Alternatively, if the hydro system owners supply the dispatcher with river chain models, rather than with half hourly station water values, the dispatchers can expect to be responsible for determining water values which are internally

---

[36] Again the issue of incentives arises since, ultimately, any dispatch can be "explained" by invoking enough constraints.

[37] These should not be confused with the long term water values specified for major storage reservoirs, though, and can not generally be determined from those by simply pro-rating, for instance.

consistent, and consistent with the dispatch, over the day. Those water values would then provide the "offer" prices required by the pricing model.

If this process has been carried out perfectly, and perfectly matches reality, the effects of all river chain constraints will be reflected properly in those prices, and there should be no discrepancies between the observed dispatch and the unconstrained dispatch which appears to be optimal at those declared water values. With no discrepancies to explain, there should be no need to invoke additional constraints in the pricing model. In reality, minor deviations from *ex post* optimality are bound to occur, but those can be handled by the methods discussed in Section 3.3.7 and Chapter 7.

In principle, a similar approach can be applied to inter-temporal linkages in the thermal system, with the possible complication of integer restrictions, as discussed below. In this case, though, it is almost certain that the dispatcher will be required to optimise the dispatch to constraints, rather than to prices, and to provide prices which are consistent with that dispatch, *ex post*. A priori, a ramping constraint could have the effect of lowering the effective marginal cost of supplying power in the period in which generation is lower, or of raising it in the period in which it is higher. Thus one approach to resolving this issue would be to internalise it within a multi-period pricing model. Alternative regimes which involve appropriate modifications to the offer prices assumed in each of the periods individually and/or a notional transfer payment between periods are discussed in Chapter 8. In the mean time, it is suggested that, where such constraints are believed to be binding, they be treated as simple upper and lower bounds on generation in the relevant periods[38]. Thus:

- If output from the generating unit in the example above was restricted because it was ramping up from a previous period or ramping down for the following period, it was effectively generating as much as it could, thus:

Nodal energy price floor =2c/kW, no ceiling

- If output from the generating unit in the example above was restricted because it was ramping down from a previous period or ramping up for the following period, it was effectively generating as little as it could, thus:

---

[38]   Such constraints will effectively remove the affected generators from consideration as possible marginal generators in the calculation of prices. This should ensure pricing consistency within periods, but not necessarily between periods.

Nodal energy price ceiling =2c/kW, no floor

3.3.5. Changed, or Unforseen, Circumstances

Uncertainty is an obvious, and pervasive, reason why dispatches are likely to appear, with hindsight, to have been sub-optimal. The dispatcher will always have to prepare for contingencies that generally do not eventuate, and will incur costs in doing so. In retrospect, those costs will most often appear not to be justified, because the contingency did not occur. Conversely, if the contingency actually occurs, the provision made for it will most likely appear less generous than that which should have been made by a dispatcher with perfect foresight.

Reserve requirements, which are discussed in a stochastic context in Chapter 9, provide a special case of this general phenomenon. There are two possible approaches to reserve pricing, or more generally, to pricing under uncertainty, reflecting two possible interpretations of the *ex post* pricing philosophy:

- A strict approach to *ex post* pricing would insist that prices be based on dispatch <u>outcomes</u>, so that those generating companies which choose, or are required, to provide reserve would receive no reward, and hence incur and opportunity cost, unless a contingency actually occurred, in which case prices would have to rise to very high levels during this contingency. For the case of node *i*, in the absence of losses, we could conclude that the price should satisfy the following conditions:

  No Contingency:

    Combined nodal energy price/opportunity cost floor = ceiling
    =2c/kW

    Nodal energy price = system marginal cost =2.2c/kW

  Contingency:

    Nodal energy price floor =2c/kW, no ceiling

    Nodal energy price = reserve price = system marginal cost
    $(\geq 2.2c/kW)$

- A less strict approach would allow prices to be based on dispatch <u>intentions</u>, so that those generating companies which choose, or are required, to provide reserve would be paid a stand-by fee, irrespective of whether a contingency actually occurs or not, and may receive little or nothing extra during a contingency. The payment of the stand-by fee effectively gives the dispatcher the right to dispatch those generating units covered by the fee. For this case we have:

Contingency or no-contingency

Stand-by free paid (no price bounds)

Combined nodal energy price/opportunity cost floor = ceiling
=2c/kW

Nodal energy price = system marginal cost = 2.2c/kW

In theory these could produce equivalent outcomes, with a low probability of very high contingency prices raising the expected reward for reserve providers in the first case to the same level as the reserve fee required in the second[39]. We argue in Chapter 9, though, that the value which both dispatcher and dispatched are likely to place on ensuring some degree of certainty in this situation make the latter a more practical approach. Hence, in part, our preference for the term "dispatch based" pricing rather than *ex post* pricing.

3.3.6. Integer Effects

An issue which has received relatively little attention in this literature (except for S. Smith (1993)) is the integer nature of unit commitment decisions and their associated costs. Strictly speaking, all of the dual analyses discussed here, and elsewhere in the literature, apply only on the assumption of a particular unit commitment pattern, and do not reflect the implications of any costs or constraints associated with unit commitment. In practice, though, these costs and constraints can be quite significant, particularly in a small system such as that in New Zealand where, for example, the removal of one 250MW unit from the dispatch can make a significant difference to the calculated spot prices. Further,

---

[39]   In both cases the rewards for reserve service must at least equal the opportunity costs of withdrawing megawatts from the energy dispatch before generators will be prepared to participate, or before such participation would be economically justified.

it will often be the case that prices calculated assuming the unit is on will suggest that it should be off, while those calculated assuming the unit is off will suggest that it should be on. In such cases there is no set of energy prices which is strictly consistent with the observed dispatch, and the dispatcher must choose the best option on the basis of overall cost but, whatever choice is made, it will appear sub-optimal when evaluated at the *ex post* prices calculated for the observed commitment/dispatch. This may also give the generating company grounds to claim compensation from the dispatcher. This issue is addressed in Chapter 10.

3.3.7. Truly Sub-Optimal Dispatch

The final possibility to be considered is that a dispatch which appears to be sub-optimal truly is. In this case there simply is no internally consistent set of prices compatible with the observed dispatch, even after all actual, or possible contingent, constraints have been accounted for. Two approaches are then possible. We could assume that the dispatcher was not responsible to strictly optimise those aspects of the dispatch which still appear non-optimal, in which case we can produce a set of prices which are consistent with the rest of the dispatch, but which may have some negative commercial implications for some parties. As mentioned above, this is the approach currently taken to reactive power and voltage pricing in New Zealand, for example. Alternatively, we could assign responsibility to the dispatcher, and then find the "best compromise" prices[40], defined as the internally consistent price set which minimises the (potential) liability of the dispatcher to compensate parties which are dispatched in an economically inconsistent manner. This provides a much more robust pricing methodology than approaches which can only produce feasible price sets if the dispatch is actually optimal, for the constraints assumed. For node *i*, the pricing problem can be expressed as follows:

Minimise the compensation payment

Combined nodal energy price/compensation payment floor = ceiling
     =2c/kW

Nodal energy price = system marginal cost = 2.2c/kW

---

[40]   Refer to Chapter 7.

3.5. Conclusion

The basic methodology of dispatch based pricing is to solve a linear programming problem which attempts to find prices which "explain" the observed dispatch. This pricing process is driven by the observation of what appears to be out of merit order dispatch, and by the search for physical constraints which explain it, rather than by an attempt to monitor or reconstruct the dispatch processes themselves. Nonetheless there is a very close relationship between the pricing problem and the original dispatch problem, and only explanations which make physical and economic sense will produce plausible prices. This approach is conceptually similar to the models previous presented, but differs in the respect that the underlying philosophy is not built purely around mathematical duality theory, but about the condition that:

> *Prices should be determined, after the fact, so as to be consistent with the observed dispatch, in the sense that they provide a rational economic explanation for it, given the price "floors" and "ceilings" expressed by the sellers and buyers in their pre-dispatch offers and bids.*

A mathematical model used to determine such prices is simply a tool to aid in determining internally consistent economic interpretations. It can only explain the examples of OOMOD which are notified to it by setting price constraints corresponding to constraints on the dispatch. Ultimately, though, the dispatcher must take responsibility for those choices, and can expect that other parties who are commercially affected may challenge these judgements.

Even when all constraints which are believed to have been binding within the period are accounted for, some discrepancies are likely to remain. These may be due to problems in interpreting the data, to constraints of a type not accounted for in the model, or which have not been properly communicated to the model, or to genuinely sub-optimal dispatch. For all of these reasons, it is recommended that an attempt be made to record dispatch intentions, at the time when decisions are made, and to use a mathematical model to arrive at as coherent an explanation of the dispatch as is possible, soon after the event.

The key advantages of dispatch based pricing are that:

- It protects the commercial interests of all parties, while allowing dispatchers, rather than being constrained by arbitrary "physical" contracting arrangements[41], to control the system in a secure and economically rational way, knowing that the economic implications of that dispatch will be consistently reflected in prices which deliver commercial and economic benefits to all parties.

- It allows the physical and economic complexity of network operation to be reflected in prices to the extent that the most sophisticated market participants deem it to be commercially worthwhile, while allowing, and giving appropriate incentives to, all parties, possibly including the dispatcher, to act in response to the best, and most up-to-date set of prices they find it worthwhile to obtain.

- It provides a single unified market place in which trade and competition are encouraged.

---

[41] Although parties which, for whatever reason, would prefer not to make their resources fully available to the system and reap the commercial benefits of so doing, may still effectively opt out of merit order dispatch by limiting physical availability, or by specifying prices which are so low (high) as to imply "must run" ("never run") status.

Chapter 4

A Dispatch Based Pricing Model[42]

4.1. Introduction

In this chapter we present a "basic" mathematical model for dispatch based pricing. This model ignores many of the dynamic effects in power systems, instead taking a "snapshot" of the system and assuming that this state applies for a full dispatch period. The pricing model developed for the New Zealand system is of this form, as are the formulations of Hogan (1991) and Schweppe et al. (1988)[43]. For the time being we assume that the dispatch is optimal, while ignoring inter-temporal, integer, and stochastic effects. These assumptions will be relaxed in later chapters.

A standard engineering tool for determining the state of a power system at a given point in time is the Optimal Power Flow (OPF) problem (Huneault and Galiona, 1991). The OPF problem involves optimising some objective function while satisfying constraints on generation, transmission and system operations in general. Like Hogan, we use the OPF as a starting point for determining dispatch based prices. In fact, the model presented in this chapter is very similar to Hogan's, though we argue that our representation of the problem is somewhat more general. We describe the key features of the OPF problem, as well as a dispatch based pricing model corresponding to it. The economic interpretation of the pricing problem, which is derived in Appendix A, is explored. A brief description of the version of this model implemented by Trans Power (NZ) Limited is also described. The model presented in this chapter provides a theoretical basis upon which discussions in later chapters can be based.

---

[42]   This chapter is based on Read and Ring (1995b).

[43]   It could be argued whether the original homeostatic proposal of Schweppe et al., which allows for pricing intervals which are of the order of seconds, is strictly a snapshot approach or not. However, for practical reasons, this model is unlikely to be applied to such short intervals in practice.

4.2. The Primal Formulation

4.2.1. The Primal Variables

The primal variables we refer to in this formulation are:

- P_{Gi} and Q_{Gi}, the active and reactive power generation levels, respectively, for node i.

- P_{Di} and Q_{Di}, the active and reactive load levels, respectively, for node i.

- $P_i = P_{Gi} - P_{Di}$ and $Q_i = Q_{Gi} - Q_{Di}$, the net nodal active and reactive power injections, respectively, for node i.

- θ_i, the phase angle at node i.

- V_i, the voltage magnitude at node i.

- x_ℓ, general variable ℓ. This variable is assumed to be a dependent variable in this chapter but is used to denote both independent and dependent variables in Chapter 7.

4.2.2. Sets Used to Distinguish Nodes

In his model, Hogan treated the nodal active and reactive power injections as his independent variables, or basis. We refer to that representation as a *PQ* representation. However, as explained by Wood and Wollenberg (1984), a more generally set of independent variables can be defined in terms of a combination of active and reactive power injections, and voltage magnitudes and phase angles. We defined the nature of nodes in terms of the variables which are independent there. Thus we define:

- PV nodes: At these nodes the net active power injection, P_i and voltage magnitude, V_i, are independent variables while the net reactive power injection, Q_i, and phase angle, θ_i, are free to take values which optimise system performance while satisfying the physical constraints on the dispatch. These nodes are typically generator nodes. Mathematically, PV nodes are associated with the set *PV*.

- PQ nodes: At these nodes P_i and Q_i are independent variables while V_i and θ_i are free to vary as required. These nodes are typically demand nodes. Mathematically, PQ nodes are associated with the set *PQ*.

- The Swing Bus: The swing bus is indicated by the "set" S or the index s. At this node V_s and θ_s are the independent variables while P_s and Q_s vary freely so as to

ensure the conservation of both active and reactive power. The value of V_s is important as it influences the level of system losses, but the value of θ_s is arbitrary as only the difference between phase angles at different nodes matters, not the actual magnitude.

As the value of the only independent phase angle, being that at the swing bus, is arbitrary we refer to this representation as a PVQ representation. As derivatives with respect to a particular variable in Hogan's PQ representation may be different from that in the PVQ representations we use a superscript "A" to indicate functions which are expressed in terms of the PVQ representation while, when appropriate, a "B" is used to indicate the PQ representation. The absence of such a superscript indicates that the function is expressed in terms of the basis formed by the nodal voltage magnitudes and phase angles (which together form voltage phasors). This later "$V\theta$" basis is the "natural" basis for expressing power flow equations. Hogan (1991) and Read, Ring, and Rosevear (1995) describe the functional forms of power injections, losses, and transmission flows in terms of the $V\theta$ basis and discuss the procedure for converting between the various bases.

In addition to the sets defined above, it is convenient to define the following combinations of sets:

- $PX = (PV \cup PQ)$: The set of all the non-swing bus nodes.

- $PXS = (PX \cup S)$: The set of all nodes.

- $PVS = (PV \cup S)$: The set of all nodes at which the reactive power injection is a dependent variable.

- $PQS = (PQ \cup S)$: The set of all nodes at which voltage magnitude is a dependent variable.

The dispatch problem described below includes some general constraints and variables. These belong to the sets:

- NN: The set of all general constraints.

- ND: The set of otherwise unspecified general dependent variables.

4.2.3. The Optimal Power Flow (OPF) Problem

An OPF formulation corresponding to our PVQ representation, in standard canonical form, is described by (4.1) to (4.14). The dual multipliers for each constraint are shown on

the right. A superscript on a vector indicates the set of nodes each vector belongs to, and hence its dimensions.

$$\underset{\boldsymbol{P_D^{PXS}},\boldsymbol{Q_D^{PXS}},\boldsymbol{P_G^{PXS}},\boldsymbol{Q_G^{PXS}},\boldsymbol{V^{PXS}},\boldsymbol{x^{ND}}}{\textbf{Minimise}} \quad C\!\left(\boldsymbol{P_G^{PXS}},\boldsymbol{Q_G^{PXS}},\boldsymbol{x^{ND}}\right) \tag{4.1}$$

subject to:

Conservation of Power

$$\sum_{i\in PXS}\!\left(P_{Gi}-P_{Di}\right)-L_P^A\!\left(\boldsymbol{P_G^{PX}}-\boldsymbol{P_D^{PX}},\boldsymbol{Q_G^{PQ}}-\boldsymbol{Q_D^{PQ}},\boldsymbol{V^{PVS}}\right)=0 \qquad :\lambda_P \tag{4.2}$$

$$\sum_{i\in PXS}\!\left(Q_{Gi}-Q_{Di}\right)-L_Q^A\!\left(\boldsymbol{P_G^{PX}}-\boldsymbol{P_D^{PX}},\boldsymbol{Q_G^{PQ}}-\boldsymbol{Q_D^{PQ}},\boldsymbol{V^{PVS}}\right)=0 \qquad :\lambda_Q \tag{4.3}$$

Dependent Reactive Power Injection at *PV* Nodes

$$-Q_i^A\!\left(\boldsymbol{P_G^{PX}}-\boldsymbol{P_D^{PX}},\boldsymbol{Q_G^{PQ}}-\boldsymbol{Q_D^{PQ}},\boldsymbol{V^{PVS}}\right)+\left(Q_{Gi}-Q_{Di}\right)=0 \qquad :\mu_{Qi} \qquad \forall i\in PV \tag{4.4}$$

Dependent Voltage Magnitude at *PQ* Nodes

$$-V_i^A\!\left(\boldsymbol{P_G^{PX}}-\boldsymbol{P_D^{PX}},\boldsymbol{Q_G^{PQ}}-\boldsymbol{Q_D^{PQ}},\boldsymbol{V^{PVS}}\right)+V_i=0 \qquad :\mu_{Vi} \qquad \forall i\in PQ \tag{4.5}$$

Generalised Constraints

$$H_h^A\!\left(\boldsymbol{P_G^{PXS}},\boldsymbol{P_D^{PXS}},\boldsymbol{Q_G^{PQS}},\boldsymbol{Q_D^{PQS}},\boldsymbol{V^{PXS}},\boldsymbol{x^{ND}}\right)=0 \qquad :\nu_h \qquad \forall h\in NN \tag{4.6}$$

Active and Reactive Generation and Voltage Bounds

$$-P_{Gi}\geq -P_{Gi}^{max} \quad :\upsilon_{Pi}^+ \qquad \forall i\in PXS \tag{4.7}$$

$$P_{Gi}\geq P_{Gi}^{min} \quad :\upsilon_{Pi}^- \qquad \forall i\in PXS \tag{4.8}$$

$$-Q_{Gi}\geq -Q_{Gi}^{max} \quad :\upsilon_{Qi}^+ \qquad \forall i\in PXS \tag{4.9}$$

$$Q_{Gi}\geq Q_{Gi}^{min} \quad :\upsilon_{Qi}^- \qquad \forall i\in PXS \tag{4.10}$$

$$-V_i\geq -V_i^{max} \quad :\upsilon_{Vi}^+ \qquad \forall i\in PXS \tag{4.11}$$

$$V_i\geq V_i^{min} \quad :\upsilon_{Vi}^- \qquad \forall i\in PXS \tag{4.12}$$

Active and Reactive Power Load Settings

$$P_{Di}=P_{Di}^{set} \quad :\beta_{Pi} \qquad \forall i\in PXS \tag{4.13}$$

$$Q_{Di}=Q_{Di}^{set} \quad :\beta_{Qi} \qquad \forall i\in PXS \tag{4.14}$$

The Independent and Dependent Variables

In vector notation, where the superscript describes the dimension, the variables in this OPF formulation are P_G^{PXS}, P_D^{PXS}, Q_G^{PXS}, Q_D^{PXS}, V^{PXS}, and x^{ND} where the independent variables are P_G^{PX}, P_D^{PXS}, Q_G^{PQ}, Q_D^{PXS}, V^{PVS} while the dependent variables are P_{Gs}, Q_G^{PVS}, V^{PQ}, and x^{ND}. We include generation variables at demand nodes, and demand variables at generator nodes, even if these will always have zero values, so as to allow us to determine the prices for these commodities if they were to be available there. Many OPF functions are only dependent on the net nodal power injections, rather than nodal generation and demand separately. Hence, even if one of either generation and demand at a node is an independent variable, the net combination of the two may still be a dependent variable. In particular the swing bus active and reactive power net injections ($P_s = P_{Gs} - P_{Ds}$ and $Q_s = Q_{Gs} - Q_{Ds}$) are dependent variables, as are the reactive power injections at *PV* nodes ($Q^{PV} = Q_G^{PV} - Q_D^{PV}$). More technical features, such as tap changing transformer settings (which are independent variables), are not considered here but have been incorporated into the pricing model employed in New Zealand (Read et al., 1995).

The Objective Function

The objective function, (4.1), represents the minimisation of the total cost associated with operating the power system[44]. This differs from Hogan's formulation in that Hogan assumed an objective of maximising total welfare. Our approach is equivalent however, as we have introduced constraints defining the observed nodal demand levels for both active and reactive power as being fixed ((4.13) and (4.14))[45]. These constraints define the price multipliers which correspond to Hogan's benefit function derivatives. Our formulation has the advantage of being more consistent with the engineering interpretation of the OPF problem.

A fuel cost is assumed for both active and reactive power generation. In minimising the value of this objective function it is necessary that the constraints which follow are satisfied.

The Power Conservation Equations

The conservation of power equations, (4.2) and (4.3), impose the requirement that total generation must equal total demand plus losses for both active and reactive power.

---

[44]   Stott et al. (1987) discuss a range of possible OPF objective functions.

[45]   Although fixed, these are still regarded as independent variables.

The swing bus must react in real time to ensure this condition is satisfied. Total system active power losses are described by the function $L_P^A\left(\boldsymbol{P_G^{PX}} - \boldsymbol{P_D^{PX}}, \boldsymbol{Q_G^{PQ}} - \boldsymbol{Q_D^{PQ}}, \boldsymbol{V}^{PVS}\right)$ while $L_Q^A\left(\boldsymbol{P_G^{PX}} - \boldsymbol{P_D^{PX}}, \boldsymbol{Q_G^{PQ}} - \boldsymbol{Q_D^{PQ}}, \boldsymbol{V}^{PVS}\right)$ describes total reactive power losses. Both these functions are represented as being dependent on the independent net nodal power injections and all the independent voltages.

The Dependent Reactive Power Injections at PV Nodes

At *PV* nodes the net injection of reactive power, $Q_i = Q_{Gi} - Q_{Di}$ $\forall i \in PV$, is defined by the independent variables. Hence (4.4) defines these injections in terms of those variables via the function $Q_{Gi}^A\left(\boldsymbol{P_G^{PX}} - \boldsymbol{P_D^{PX}}, \boldsymbol{Q_G^{PQ}} - \boldsymbol{Q_D^{PQ}}, \boldsymbol{V}^{PVS}\right)$ $\forall i \in PV$. The levels of load and generation at these nodes are not individually fixed by these constraints but their difference is.

The Dependent Voltage Magnitudes at PQ Nodes

Similarly, at *PQ* nodes the voltage terms, V_i, $\forall i \in PQ$ are also dependent on the independent net nodal injections and voltage variables. Equation (4.5) represents this dependence using the function $V_i^A\left(\boldsymbol{P_G^{PX}} - \boldsymbol{P_D^{PX}}, \boldsymbol{Q_G^{PQ}} - \boldsymbol{Q_D^{PQ}}, \boldsymbol{V}^{PVS}\right)$.

The Generalised Constraints

Constraint (4.6) is a general constraint which in later chapters will be used to examine the effects on prices of several specific constraints. The function $H_h^A\left(\boldsymbol{P_G^{PXS}}, \boldsymbol{P_D^{PXS}}, \boldsymbol{Q_G^{PQS}}, \boldsymbol{Q_D^{PQS}}, \boldsymbol{V}^{PXS}, \boldsymbol{x}^{ND}\right)$ is dependent on all the variables so as to allow any relationship between these variables to be represented, and consequently can have the effect of reducing the size of the set of independent variables. This constraint can also define both the functional form of, and the bounds on, the general dependent variables $\boldsymbol{x}^{ND}$.

Active and Reactive Generation and Voltage Bounds

The bounds (4.7) to (4.12) place constant upper (indicated by *max*) and lower bounds (indicated by *min*) on values that each of the variables in the model can take. By setting the upper and lower bounds to be identical we can fix the values of variables externally to the OPF problem.

The Load Setting Constraints

The load setting constraints, (4.13) and (4.14) fix the levels of active and reactive power demand at all nodes, i.e., the demands are set externally to the dispatch problem.

For an optimal observed setting of these values, therefore, the shadow price on each of these bounds should equal the price determined by a welfare maximising objective function when demand is not fixed. Note that fixing reactive power demand at a node still allows the total net reactive power injection at that node to vary if reactive generation can be varied. Demand has been fixed to simplify the discussion, but could be optimised like generation if required.

4.2.4. Convexity of the Primal OPF

We assume that the primal OPF is, at least, locally convex about an observed solution. While the equality constraints in (4.2) to (4.6) make them appear non-convex, this limitation can be overcome in most situations. Firstly, we explicitly assume that (4.6) is convex, though allow it to have non-differentiable points. Secondly, the OPF is solved with a fixed set of generating units dispatched, and with at least one being a slack variable, so the energy conservation constraints, (4.2) and (4.3), can be viewed as inequalities (Glavitsh, 1992). There is no certainty that (4.4) and (4.5) are convex, but Hogan, Read, and Ring (1995) observe that local non-convexity only appears to be an issue in lightly loaded power systems, and that, in this event, the corresponding power flow may be only locally optimal. While this may have some revenue implications for the grid owner, which collects economic rents from the dispatch, these can always be offset by separate fixed cost recovery mechanisms. We adopt the conjecture of Hogan et al. that under normal operating conditions the impact on the spot market of non-convexities will be minimal.

4.3. The Dual Pricing Formulation

4.3.1. The Mathematical Form of the Dual

We can form a pricing model by linearising the OPF of (4.1) to (4.14) and forming its dual. This dual, which is derived in Appendix B, is described by equations (4.15) to (4.24). The definitions of all the new terms are given in the discussion which follows.

$$\underset{\substack{\boldsymbol{\nu}^{NN}, \boldsymbol{v}_P^{+PXS}, \boldsymbol{v}_Q^{+PXS}, \boldsymbol{v}_V^{+PXS} \geq \mathbf{0} \\ \boldsymbol{v}_P^{-PXS}, \boldsymbol{v}_Q^{-PXS}, \boldsymbol{v}_V^{-PXS} \geq \mathbf{0} \\ \lambda_P, \lambda_Q, \boldsymbol{\mu}_Q^{PV}, \boldsymbol{\mu}_V^{PQ}, \boldsymbol{\beta}_P^{PXS}, \boldsymbol{\beta}_Q^{PXS}}}{\textbf{Minimise}} \qquad \text{Arbitrary Objective Function} \tag{4.15}$$

subject to :

$$\beta_{Pi} = \lambda_P \left(1 - \frac{\partial L_P^A}{\partial^A P_i}\right) - \lambda_Q \frac{\partial L_Q^A}{\partial^A P_i} - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A P_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A P_i} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Di}} \right| \qquad \forall i \in PXS \tag{4.16}$$

$$\beta_{Qi} = -\lambda_P \frac{\partial L_P^A}{\partial^A Q_i} + \lambda_Q \left(1 - \frac{\partial L_Q^A}{\partial^A Q_i}\right) - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A Q_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A Q_i} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Di}} \right| \qquad \forall i \in PQS \tag{4.17}$$

$$\beta_{Vi} = -\lambda_P \frac{\partial L_P^A}{\partial^A V_i} - \lambda_Q \frac{\partial L_Q^A}{\partial^A V_i} - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A V_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A V_i} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A V_i} \right| \qquad \forall i \in PVS \tag{4.18}$$

$$\mu_{Qi} = \beta_{Qi} - \lambda_Q + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Di}} \right| \qquad \forall i \in PV \tag{4.19}$$

$$\mu_{Vi} = \beta_{Vi} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A V_i} \right| \qquad \forall i \in PQ \tag{4.20}$$

$$c_{Pi}^- - \langle v_{Pi}^- \rangle + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Gi}^-} \right| \leq \beta_{Pi} + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Di}} \right| \leq c_{Pi}^+ + \langle v_{Pi}^+ \rangle - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Gi}^+} \right| \qquad \forall i \in PXS \tag{4.21}$$

$$c_{Qi}^- - \langle v_{Qi}^- \rangle + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Gi}^-} \right| \leq \beta_{Qi} + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Di}} \right| \leq c_{Qi}^+ + \langle v_{Qi}^+ \rangle - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Gi}^+} \right| \qquad \forall i \in PXS \tag{4.22}$$

$$\beta_{Vi} = \langle v_{Vi}^+ \rangle - \langle v_{Vi}^- \rangle \qquad \forall i \in PXS \tag{4.23}$$

$$c_{x\ell}^- + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A x_\ell^-} \right| \leq 0 \leq c_{x\ell}^+ - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A x_\ell^+} \right| \qquad \forall \ell \in ND \tag{4.24}$$

4.3.2 The Pricing Objective Function

The pricing objective function (4.15) is stated as having an arbitrary form[46] because, by Theorem A.2. in Appendix A, if the primal dispatch solution is already known, and if it is optimal, then the dual objective function is arbitrary as long as the complementary slackness conditions are imposed directly. The latter condition is enforced in this dual by the $\langle z \rangle$ notation which states that the term z only appears in the model if the primal constraint it corresponds to is binding. Ring, Read, and Drayton (1993) conclude that the only role that the dual objective function serves is to determine how the total wealth, which is fixed given the observed primal solution, is distributed between the market players, and hence that the choice of objective is ultimately a policy issue. If the dual solution is not degenerate, which can be expected to be the normal situation, then the arbitrary pricing objective function plays no role in determining the pricing solution.

Some possible pricing objective functions are discussed by Read and Ring (1995b). They discuss objective functions which either minimise or maximise the rent collected by the grid owner. While there is no obvious "best" choice of objective function, the former approach maximises the cumulative rent collected by the market players, while the latter approach provides a non-distortionary way of reducing the additional fixed charges required if all transmission network costs are to be recovered.

4.3.3. Active Power Pricing Relationships

Equation (4.16) describes the price for active power <u>demand</u> at all nodes. The swing bus net active power injection ($P_s = P_{Gs} - P_{Ds}$) is a dependent variable so all the derivatives in (4.16) with respect to this net injection are zero. Consequently, the only non-zero derivative for the swing bus is that of the general constraint, i.e.:

$$\beta_{Ps} = \lambda_P - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Ds}} \right\rangle \qquad (4.25)$$

That is, the active power price at the swing bus is just the "unit lambda", λ_P, which, as the shadow price on the active power energy conservation constraint, describes the marginal value of active power at the swing bus[47], modified by the impact of the swing bus on any

---

[46]   And hence the choice of whether it is maximised or minimised is arbitrary.

[47]   In practice it is possible to treat any node as a "swing bus" for pricing purposes but the price at that node must be equated to that nodes "unit lambda". The price at each node which is a marginal source of active power is then just the pricing swing bus "unit lambda" modified by the marginal cost of losses and constraints involved in (notionally) transmitting one more unit of power from the

general constraints. No other terms appear because changes in (notional) demand at the swing bus are assumed to be made up at the swing bus and hence have no impact on the network. If there were no binding general constraints then λ_P would be the price at the swing bus.

The active power price at the non-swing bus nodes can be re-expressed as:

$$\beta_{Pi} = \lambda_P - \left(\lambda_P \frac{\partial L_P^A}{\partial^A P_i} + \lambda_Q \frac{\partial L_Q^A}{\partial^A P_i} \right) - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A P_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A P_i} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Di}} \right\rangle$$

$$\forall i \in PX \quad (4.26)$$

We can conclude that a unit increase in active power demand at each node results in five types of costs:

- The first term on the left is just the cost of actually supplying that extra unit of active power at the swing bus, whether by generating it there or by importing it from the marginal node(s).

- The second term, $-\left(\lambda_P \frac{\partial L_P^A}{\partial^A P_i} + \lambda_Q \frac{\partial L_Q^A}{\partial^A P_i} \right)$, reflects marginal cost of the additional power required at the swing bus to make up for the losses (both active and reactive) resulting from the transmission of that power.

- The third term, $-\sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A P_i}$, describes the cost (or benefit) of increasing (or decreasing) pressure on binding reactive power bounds at *PV* nodes.

- The fourth term, $-\sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A P_i}$, describes the cost (or benefit) of increasing (or decreasing) pressure on binding voltage bounds at *PQ* nodes.

- The fifth term, $-\sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Di}} \right\rangle$, describes the cost (of benefit) or increasing (or decreasing) pressure on a binding general constraint which is functionally dependent on active power demand at that node.

The last three cost components are discussed further below.

---

pricing swing bus to the marginal node. By definition, the resulting price at a unique marginal node must be the so called "system lambda".

Equation set (4.21) places bounds on the feasible values that the price for active power demand can take at all nodes. In the absence of binding general constraint terms these become:

$$c_{Pi}^- - \left\langle v_{Pi}^- \right\rangle \leq \beta_{Pi} \leq c_{Pi}^+ + \left\langle v_{Pi}^+ \right\rangle \qquad\qquad \forall i \in PXS \quad (4.27)$$

Here the terms c_{Pi}^- and c_{Pi}^+ are, respectively, the marginal costs of the last unit of active power generated at node i and the next unit of active power that could be generated there[48].

If no generating unit bounds are binding at node i then $v_{Pi}^- = v_{Pi}^+ = 0$ and, before solving the pricing problem, we can set $\left\langle v_{Pi}^- \right\rangle = \left\langle v_{Pi}^+ \right\rangle = 0$. Thus, at node i, we have:

$$c_{Pi}^- \leq \beta_{Pi} \leq c_{Pi}^+ \qquad\qquad (4.28)$$

That is, the price paid for active power at node i must be consistent with the observed cost of generation at that node. If this were not the case then either a generating unit would be producing its last unit of generation at a loss ($\beta_{Pi} \leq c_{Pi}^-$) or failing to produce a profitable additional unit of generation ($c_{Pi}^+ \leq \beta_{Pi}$). If we were to have $c_{Pi} = c_{Pi}^- = c_{Pi}^+$, with no generating unit bounds binding at node i, then the price at that node must exactly equal the generation cost. Note that if a generating unit is partially loaded it is not necessarily marginal unless $c_{Pi}^- = c_{Pi}^+$. This is because, if $c_{Pi}^- < c_{Pi}^+$, then there might be partially loaded generating units which could meet an increase in load more cheaply, or save more money by reducing output in response to a decrease in load.

We now consider what happens when active power generation at node i is constrained. Duality theory dictates that in this situation $v_{Pi}^+ \geq 0$ or $v_{Pi}^- \geq 0$, depending on which bound is binding. The value of v_{Pi}^+ reflects the change in the cost of running the system if we relax a binding upper bound on generation. If the upper bound were relaxed by one unit the cost of operating the system would fall, as less power would be required from more expensive sources. Thus the change in the cost of operating the system is negative. However, we have defined the primal constraint as $-P_{Gi}^+ \geq -P_{Gi}^{max} + P_{Gi}^*$, so v_{Pi}^+ actually describes the negative value of a change in the upper bound. Hence the value of v_{Pi}^+ must be positive to reflect this decrease in the system operating cost. If the upper bound were not limiting the generating unit's output then we would have $v_{Pi}^+ = 0$.

Similarly, the value of v_{Pi}^- reflects the value of changing a generating unit's lower bound. We have defined the primal constraint as $-P_{Gi}^- \geq P_{Gi}^{min} - P_{Gi}^*$ so v_{Pi}^- describes the

---

[48]   These cost terms could be used to represent demand side bids, if required.

positive value of a change in the lower bound. Increasing a binding lower bound on generation means that we are forced to generate power which we do not want to produce, presumably because this generation is not profitable. Hence the change in system cost from increasing the bound is positive, so v_{Pi}^- is positive.

From (2.27) we see that with either $v_{Pi}^- > 0$ (generation at lower bound) or $v_{Pi}^+ > 0$ (generation at upper bound) β_{Pi} can take values outside the range defined by the generation costs. For example, a generating unit may have its output limited by its upper bound, implying $v_{Pi}^+ > 0$. The marginal fuel cost of this generating unit is $c_{Pi}^+ = \$10/\text{MW}$, but consistency with other prices requires $\beta_{Pi} = \$15/\text{MW}$. We can find consistent prices for this situation by setting $v_{Pi}^+ = \$5/\text{MW}$.

The inclusion of the general constraints further modifies the range of feasible prices. These modifications are discussed in the context of specific constraints described in later chapters.

4.3.4. Reactive Power Pricing Relationships - PQS Nodes

The prices of reactive power at *PQS* nodes are similar to those for active power. Again, the swing bus price has a simpler form than the other equations in (4.17) as the net injection of reactive power is again a dependent variable:

$$\beta_{Qs} = \lambda_Q - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Ds}} \right\rangle$$

The price at each of the *PQ* nodes, where the net injections of reactive power are independent variables, can be re-expressed as:

$$\beta_{Qi} = \lambda_Q - \left(\lambda_P \frac{\partial L_P^A}{\partial^A Q_i} + \lambda_Q \frac{\partial L_Q^A}{\partial^A Q_i} \right) - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A Q_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A Q_i} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Di}} \right\rangle$$

$$\forall i \in PQ \quad (4.29)$$

As for the active power prices, we can group the costs associated with a unit increase in demand for reactive power at one of the *PQ* nodes as follows:

- The first term on the left, λ_Q, is the reactive power analog of the active power unit lambda, and is just the cost of actually supplying that extra unit of reactive power at the swing bus, whether by generating it there or by importing it.

- The second term, $-\left(\lambda_P \dfrac{\partial L_P^A}{\partial^A Q_i} + \lambda_Q \dfrac{\partial L_Q^A}{\partial^A Q_i}\right)$ reflects marginal cost of the additional

  power required at the swing bus to make up for the losses (both active and reactive) resulting from the transmission of that power.

- The third term, $-\displaystyle\sum_{n \in PV} \mu_{Qn} \dfrac{\partial Q_n^A}{\partial^A Q_i}$, describes the cost (or benefit) of increasing (or

  decreasing) pressure on binding reactive power bounds at *PV* nodes.

- The fourth term, $-\displaystyle\sum_{n \in PQ} \mu_{Vn} \dfrac{\partial V_n^A}{\partial^A Q_i}$, describes the cost (or benefit) of increasing (or

  decreasing) pressure on binding voltage bounds at *PQ* nodes.

- The fifth term, $-\displaystyle\sum_{h \in NN} \langle v_h \rangle \left\langle \dfrac{\partial H_h^A}{\partial^A Q_{Di}} \right\rangle$, describes the cost (of benefit) or increasing (or

  decreasing) pressure on a binding general constraint which is functionally dependent on reactive power demand at that node.

The bounds on the reactive power prices are defined by (4.22). The explanation of these bounds is entirely analogous to the explanation for the active power price bounds. However, it may occur that reactive power can be produced at no cost. In this situation, and in the absence of general constraints, (4.22) can be written as:

$$\beta_{Qi} = \langle v_{Qi}^+ \rangle - \langle v_{Qi}^- \rangle \tag{4.30}$$

While this form does not directly follow from (4.22), the price must either equal the multiplier on the upper bound or the negative of the multiplier on the lower bound. If no bounds are binding on reactive power generation then the price is zero. Alternatively, if reactive bounds are binding, then the price will take what ever value is consistent with the dispatch. In particular, this latter case applies if reactive power injections are set externally to the OPF, rather than being optimised.

4.3.5. Reactive Power Pricing Relationships - PV Nodes

At *PV* nodes reactive power injections are dependent variables, much like the swing bus reactive power injection. By re-arranging (4.19), and assuming no general constraints are binding, the reactive power price at these nodes can be described as:

$$\beta_{Qi} = \lambda_Q + \mu_{Qi} \qquad\qquad \forall i \in PV \tag{4.31}$$

Since μ_{Qi}, the multiplier on the OPF constraint defining the dependence of the reactive injections at *PV* nodes on the independent variables, may be positive or negative, this

equation merely allows β_{Qi}, to take on whatever value is required, independent of λ_Q. Obviously, if $\mu_{Qi} = 0$ for some subset of *PV* nodes, then we have $\beta_{Qi} = \lambda_Q$, reflecting the fact that each of these nodes plays the same role as the swing bus in meeting increases (or decreases) in reactive power demand and losses.

However, β_{Qi} may have to differ from β_{Qs} for several reasons. First, the bounds imposed by the reactive power generation costs at these nodes, described by (4.22), may force a difference. Second, as for *PQ* nodes, v_{Qi}^+ and v_{Qi}^- may allow β_{Qi} to vary if reactive power generation is constrained at this node. Finally, the actual value of β_{Qi} will be driven, subject to the above considerations, by the need to obtain consistent prices for the other commodities for which it appears in price equations or, if there is any choice in the matter, by the objective function.

4.3.6. Voltage Pricing Relationships - PVS Nodes

The equations describing active and reactive power injections, line flows, and losses as functions of nodal voltage magnitudes and phase angles (i.e., the $V\theta$ basis described above) reveal a quadratic dependency of the former terms on voltage (Read, Ring, and Rosevear, 1995). In practice, this gives rise to a non-trivial relationship between voltage settings and the cost of running a power system. Bohn, Caramanis and Schweppe (1984) noted this effect and observed that in principal prices could be applied to voltages. If such prices could be used as a means of encouraging modifications in voltage levels than significant savings could potentially be made. Boshier (1974) reinforces this argument by concluding that the optimisation of voltage levels in New Zealand could, at that time, reduce generation costs by 0.25% to 0.75% . On the demand side, Dias and El-Hawary (1990) observe that when active and reactive loads are modelled as functions of voltage the total fuel cost of a system could be reduced by keeping voltages as low as possible within allowable limits. The significance of these result is that voltage pricing information may be valuable in achieving optimal resource usage in power systems.

Unlike nodal power injections, voltage does not consist of a "demand" component and a "generation" component. Mathematically, therefore, a voltage price can either be modelled analogously to generation, making a positive "net injection" of voltage, or to demand, making a negative "net injection". It seems preferable, however, to adopt a definition corresponding to pricing the "net demand" for voltage, in a manner consistent with prices determined for active and reactive power net injections. Thus we introduce a voltage price analogously to active and reactive power prices, as shown in (4.18). This can be re-expressed as:

$$\beta_{Vi}= 0-\left(\lambda_P \frac{\partial L_P^A}{\partial^A V_i}+\lambda_Q \frac{\partial L_Q^A}{\partial^A V_i}\right)-\sum_{n\in PV}\mu_{Qn}\frac{\partial Q_n^A}{\partial^A V_i}-\sum_{n\in PQ}\mu_{Vn}\frac{\partial V_n^A}{\partial^A V_i}-\sum_{h\in NN}\langle v_h\rangle\left\langle\frac{\partial H_h^A}{\partial^A V_i}\right\rangle$$

$$\forall i\in PVS \quad (4.32)$$

The components of this equation can be described as follows:

- The zero immediately to the right of the equality sign indicates that there is no direct energy cost associated with additional units of voltage.

- The second term, $-\left(\lambda_P \dfrac{\partial L_P^A}{\partial^A V_i}+\lambda_Q \dfrac{\partial L_Q^A}{\partial^A V_i}\right)$ reflects marginal cost of the additional power required at the swing bus to make up for the losses (both active and reactive) resulting from a change in voltage.

- The third term, $-\sum_{n\in PV}\mu_{Qn}\dfrac{\partial Q_n^A}{\partial^A V_i}$, describes the cost (or benefit) of increasing (or decreasing) pressure on binding reactive power bounds at *PV* nodes.

- The fourth term, $-\sum_{n\in PQ}\mu_{Vn}\dfrac{\partial V_n^A}{\partial^A V_i}$, describes the cost (or benefit) of increasing (or decreasing) pressure on binding voltage bounds at *PQ* nodes.

- The fifth term, $-\sum_{h\in NN}\langle v_h\rangle\left\langle\dfrac{\partial H_h^A}{\partial^A V_i}\right\rangle$, describes the cost (of benefit) or increasing (or decreasing) pressure on a binding general constraint which is functionally dependent on voltage at that node.

The prices defined by (4.32) are constrained by those voltage price bounds of (4.23) which correspond to *PVS* nodes. That is:

$$\beta_{Vi}=\langle v_{Vi}^+\rangle-\langle v_{Vi}^-\rangle \qquad\qquad \forall i\in PVS \quad (4.33)$$

It may be seen that this equation has a similar form to the bounds on β_{Pi} and β_{Qi} (i.e. (4.21) and (4.22)), but without any marginal cost terms (as none are assumed) or general constraint derivatives (which only appear in the voltage "demand" derivative of (4.18)). If voltage has been freely optimised, with neither upper or lower bounds binding, then $v_{Vi}^- = v_{Vi}^+ = 0$ and we require that:

$$\beta_{Vi} = 0 \qquad\qquad \forall i\in PVS \quad (4.34)$$

Otherwise, v_{Vi}^+ and v_{Vi}^- allow β_{Vi} to be non-zero, just as for the prices for costless reactive power . In particular, if V_i has been set externally to the dispatch, then β_{Vi} is unrestricted in sign.

4.3.7. Voltage Pricing Relationships - PQ Nodes

The arguments for calculating voltage prices at *PVS* nodes apply equally to *PQ* nodes, though the pricing relationships, described by (4.20), are much simpler. In the absence of binding general constraints these relationship simplify to:

$$\mu_{Vi} = \beta_{Vi} \qquad\qquad \forall i \in PQ \quad (4.35)$$

Since μ_{Vi}, the multiplier on the constraint defining the dependent voltages, is unconstrained in sign, (4.35) simply states that the voltage price is free to take any value, subject only to those bounds in (4.23) which apply to *PV* nodes. That is:

$$\beta_{Vi} = \langle v_{Vi}^+ \rangle - \langle v_{Vi}^- \rangle \qquad\qquad \forall i \in PQ \quad (4.36)$$

Thus, in the absence of binding general constraints, the value associated with constrained, dependent voltage is completely defined by the dual variables corresponding to the upper and lower bounds on these terms. As for *PVS* nodes, these prices will be zero if no voltage bounds are binding.

4.3.8. Pricing for the General Dependent Variables

Equation (4.24) describes the price bounds on a general dependent variable, x_ℓ, where $c_{x\ell}^-$ and $c_{x\ell}^+$ are, respectively, the costs of the last and next units of variable x_ℓ. This price bound may appear to have a fundamentally different form to those for active and reactive power, but this is just an artefact of its generality. For instance, some of the primal general constraints may include upper and lower bounds on the x_ℓ. If this is the case then we could transform those terms into a price for the x_ℓ. For example, for a given variable x_ℓ we could separate out two such terms from the sum of general constraint derivatives to give:

$$\begin{aligned}
V_1 \frac{\partial H_1^A}{\partial^A x_\ell^-} &= -v_{x\ell}^- \\
V_2 \frac{\partial H_2^A}{\partial^A x_\ell^+} &= -v_{x\ell}^+
\end{aligned} \qquad (4.37)$$

We might also, for example, have a term satisfying:

$$V_3 \frac{\partial H_3^A}{\partial^A x_\ell^-} = V_3 \frac{\partial H_3^A}{\partial^A x_\ell^+} = \alpha_{x\ell} \qquad (4.38)$$

Here α_{xl} is an arbitrary dual multiplier. With these expressions we could re-express the price bounds associated with this x_ℓ as:

$$c_{x\ell}^- - \left\langle v_{x\ell}^- \right\rangle + \sum_{h \in (NN-2)} \left\langle v_h \right\rangle \left\langle \frac{\partial H_h^A}{\partial^A x_\ell^-} \right\rangle \leq \alpha_{x\ell} \leq c_{x\ell}^+ + \left\langle v_{x\ell}^+ \right\rangle - \sum_{h \in (NN-2)} \left\langle v_h \right\rangle \left\langle \frac{\partial H_h^A}{\partial^A x_\ell^+} \right\rangle \qquad (4.39)$$

This price bound is analogous to those for active and reactive power prices.

Equation (4.24) can also be simplified, for a given x_ℓ, if all the general constraints are continuous and differentiable with respect to x_ℓ, and if no marginal cost appears in the primal objective function for x_ℓ. In this instance (4.24) reduces to:

$$\sum_{h \in NN} \left\langle v_h \right\rangle \left\langle \frac{\partial H_h^A}{\partial^A x_\ell} \right\rangle = 0 \qquad (4.40)$$

4.4. An Alternative Pricing Model Form

4.4.1. Derivation of Alternative Model

The pricing model presented in (4.15) to (4.24), like past models, involves a great deal of complexity which obscures much of the underlying structure of the problem. Fortunately, it is possible to transform this result to remove much of this complexity. For simplicity, we initially assume that none of the general constraints are binding, allowing us to eliminate the corresponding terms from the dual. The resulting formulation is shown in (4.41) to (4.49).

$$\underset{\substack{\boldsymbol{v}_P^{+PXS},\,\boldsymbol{v}_Q^{+PXS},\,\boldsymbol{v}_V^{+PXS}\geq 0 \\ \boldsymbol{v}_P^{-PXS},\,\boldsymbol{v}_Q^{-PXS},\,\boldsymbol{v}_V^{-PXS}\geq 0 \\ \lambda_P,\lambda_Q,\mu_Q^{PV},\mu_V^{PQ},\beta_P^{PXS},\beta_Q^{PXS}}}{\textbf{Minimise}} \qquad \text{Arbitrary Objective Function} \qquad (4.41)$$

subject to:

$$\beta_{Pi} = \lambda_P\left(1 - \frac{\partial L_P^A}{\partial^A P_i}\right) - \lambda_Q\frac{\partial L_Q^A}{\partial^A P_i} - \sum_{n\in PV}\mu_{Qn}\frac{\partial Q_n^A}{\partial^A P_i} - \sum_{n\in PQ}\mu_{Vn}\frac{\partial V_n^A}{\partial^A P_i} \qquad \forall i\in PXS \quad (4.42)$$

$$\beta_{Qi} = -\lambda_P\frac{\partial L_P^A}{\partial^A Q_i} + \lambda_Q\left(1 - \frac{\partial L_Q^A}{\partial^A Q_i}\right) - \sum_{n\in PV}\mu_{Qn}\frac{\partial Q_n^A}{\partial^A Q_i} - \sum_{n\in PQ}\mu_{Vn}\frac{\partial V_n^A}{\partial^A Q_i} \qquad \forall i\in PQS \quad (4.43)$$

$$\beta_{Vi} = -\lambda_P\frac{\partial L_P^A}{\partial^A V_i} - \lambda_Q\frac{\partial L_Q^A}{\partial^A V_i} - \sum_{n\in PV}\mu_{Qn}\frac{\partial Q_n^A}{\partial^A V_i} - \sum_{n\in PQ}\mu_{Vn}\frac{\partial V_n^A}{\partial^A V_i} \qquad \forall i\in PVS \quad (4.44)$$

$$\mu_{Qi} = \beta_{Qi} - \lambda_Q \qquad\qquad\qquad\qquad\qquad \forall i\in PV \quad (4.45)$$

$$\mu_{Vi} = \beta_{Vi} \qquad\qquad\qquad\qquad\qquad\qquad \forall i\in PQ \quad (4.46)$$

$$c_{Pi}^- - \left\langle v_{Pi}^-\right\rangle \leq \beta_{Pi}\leq c_{Pi}^+ + \left\langle v_{Pi}^+\right\rangle \qquad\qquad \forall i\in PXS \quad (4.47)$$

$$c_{Qi}^- - \left\langle v_{Qi}^-\right\rangle \leq \beta_{Qi}\leq c_{Qi}^+ + \left\langle v_{Qi}^+\right\rangle \qquad\qquad \forall i\in PXS \quad (4.48)$$

$$\beta_{Vi} = \left\langle v_{Vi}^+\right\rangle - \left\langle v_{Vi}^-\right\rangle \qquad\qquad\qquad \forall i\in PXS \quad (4.49)$$

We can simplify this result by substituting so as to eliminate the μ_{Qi} and μ_{Vi} terms and by grouping the resulting terms which are multiplied by λ_P and λ_Q. This result can be further simplified by observing that the OPF power energy conservation constraints can be presented in the form:

$$P_s = -\sum_{i\in PX}P_i + L_P^A\left(\boldsymbol{P}_G^{PX} - \boldsymbol{P}_D^{PX}, \boldsymbol{Q}_G^{PQ} - \boldsymbol{Q}_D^{PQ}, \boldsymbol{V}^{PVS}\right) \qquad (4.50)$$

$$Q_s = -\sum_{i\in PQ}Q_{Gi} - \sum_{i\in PV}Q_{Gi} + L_Q^A\left(\boldsymbol{P}_G^{PX} - \boldsymbol{P}_D^{PX}, \boldsymbol{Q}_G^{PQ} - \boldsymbol{Q}_D^{PQ}, \boldsymbol{V}^{PVS}\right) \qquad (4.51)$$

When determining the active power price at node i we assume a unit change in injection there, while keeping all other independent variables fixed. The only things that can respond to this change are the dependent variables, namely the swing bus injections, active and reactive losses, and the dependent reactive power injections at *PV* nodes and

dependent voltages at *PQ* nodes. This means that the derivatives of the equations defining the swing bus active and reactive power injections must be[49]:

$$\frac{\partial P_s^A}{\partial^A P_i} = -\left(\frac{\partial P_i^A}{\partial^A P_i} - \frac{\partial L_P^A}{\partial^A P_i}\right) = -\left(1 - \frac{\partial L_P^A}{\partial^A P_i}\right) \qquad \forall i \in PX \qquad (4.52)$$

$$\frac{\partial Q_s^A}{\partial^A P_i} = \left(\frac{\partial L_Q^A}{\partial^A P_i} - \sum_{n \in PV} \frac{\partial Q_n^A}{\partial^A P_i}\right) \qquad \forall i \in PX \qquad (4.53)$$

$$\frac{\partial P_s^A}{\partial^A Q_i} = \frac{\partial L_P^A}{\partial^A Q_i} \qquad \forall i \in PQ \qquad (4.54)$$

$$\frac{\partial Q_s^A}{\partial^A Q_i} = -\left(1 - \frac{\partial L_Q^A}{\partial^A Q_i} + \sum_{n \in PV} \frac{\partial Q_n^A}{\partial^A Q_i}\right) \qquad \forall i \in PQ \qquad (4.55)$$

$$\frac{\partial P_s^A}{\partial^A V_i} = \frac{\partial L_P^A}{\partial^A V_i} \qquad \forall i \in PVS \qquad (4.56)$$

$$\frac{\partial Q_s^A}{\partial^A V_i} = \left(\frac{\partial L_Q^A}{\partial^A V_i} - \sum_{n \in PV} \frac{\partial Q_n^A}{\partial^A V_i}\right) \qquad \forall i \in PVS \qquad (4.57)$$

Further, since all derivatives with respect to the swing bus injections must be zero, we can conclude from (4.42) and (4.43) that the swing bus active and reactive power prices are given by:

$$c_{Qi}^- - \left\langle v_{Qi}^- \right\rangle \qquad (4.58)$$

$$\beta_{Qs} = \lambda_Q \qquad (4.59)$$

Using these relationships we can re-express the pricing problem as shown in (4.60) to (4.66).

---

[49]   There is no derivative with respect to the swing bus active power injection itself as, being a dependent variable, that derivative is zero.

$$\underset{\substack{\boldsymbol{v}_P^+, \boldsymbol{v}_Q^+, \boldsymbol{v}_V^+ \geq 0 \\ \boldsymbol{v}_P^-, \boldsymbol{v}_Q^-, \boldsymbol{v}_V^- \geq 0 \\ \boldsymbol{\beta}_P, \boldsymbol{\beta}_Q, \boldsymbol{\beta}_V}}{\textbf{Minimise}} \quad \text{Arbitrary Objective Function} \qquad\qquad (4.60)$$

subject to:

$$\beta_{Pi} = -\beta_{Ps}\frac{\partial P_s^A}{\partial^A P_i} - \sum_{n \in PVS}\beta_{Qn}\frac{\partial Q_n^A}{\partial^A P_i} - \sum_{n \in PQ}\beta_{Vn}\frac{\partial V_n^A}{\partial^A P_i} \qquad \forall i \in PX \quad (4.61)$$

$$\beta_{Qi} = -\beta_{Ps}\frac{\partial P_s^A}{\partial^A Q_i} - \sum_{n \in PVS}\beta_{Qn}\frac{\partial Q_n^A}{\partial^A Q_i} - \sum_{n \in PQ}\beta_{Vn}\frac{\partial V_n^A}{\partial^A Q_i} \qquad \forall i \in PQ \quad (4.62)$$

$$\beta_{Vi} = -\beta_{Ps}\frac{\partial P_s^A}{\partial^A V_i} - \sum_{n \in PVS}\beta_{Qn}\frac{\partial Q_n^A}{\partial^A V_i} - \sum_{n \in PQ}\beta_{Vn}\frac{\partial V_n^A}{\partial^A V_i} \qquad \forall i \in PVS \quad (4.63)$$

$$c_{Pi}^- - \left\langle v_{Pi}^- \right\rangle \leq \beta_{Pi} \leq c_{Pi}^+ + \left\langle v_{Pi}^+ \right\rangle \qquad \forall i \in PXS \quad (4.64)$$

$$c_{Qi}^- - \left\langle v_{Qi}^- \right\rangle \leq \beta_{Qi} \leq c_{Qi}^+ + \left\langle v_{Qi}^+ \right\rangle \qquad \forall i \in PXS \quad (4.65)$$

$$\beta_{Vi} = \left\langle v_{Vi}^+ \right\rangle - \left\langle v_{Vi}^- \right\rangle \qquad \forall i \in PXS \quad (4.66)$$

The interpretation and significance of these equations are explored in the remainder of this section.

4.4.2. The Swing Bus

In this form the role of the swing bus reactive power generation is revealed to be no different from that of any other node at which the reactive power injection is a dependent variable. It is, though, the only node at which active power generation can vary in response to a change in the independent variables. We note that active power prices should be invariant to the choice of swing bus provided that the system is constrained in an identical manner in each case. In reality, a change in power injection at node *i* will be matched by changes at some marginal node(s), rather than at the swing bus. Conceptually, we can view the swing bus as responding to the change in injection at node *i*, and then responding independently to the change at the marginal node(s), the net effect being to produce no change in injection at the swing bus (as required by power conservation), making the swing bus an arbitrary reference point at which a hypothetical transaction occurs. Intuitively, the same result should apply to reactive power provided that the reactive power swing bus is chosen from among those nodes at which the reactive power injection is free to vary[50].

---

[50]   This argument rests, though, on the assumption that Q_i will react to changes in the same way when *i* is a dependent, non-swing bus, injection (i.e. *i* is a *PV* node) as it does when *i* is the swing bus (i.e. a *Vθ* node) and/or that the optimality conditions of the problem ensure that the net impact of reactive changes, at nodes other than the two directly involved, will cancel out to give the same cost, irrespective of any variation in the way those changes are apportioned between nodes. These

4.4.3. Relationship Between Primal Constraints and Dual Variables

Standard optimisation theory dictates that each additional dual constraint makes it potentially more difficult to find a feasible solution to the dual, that is to find a set of prices which make all parties satisfied. Each primal constraint, on the other hand, introduces a new dual variable, giving the dual an extra degree of freedom and hence increases the range of feasible solutions, making the dual easier to solve. Basically, the existence of primal constraints makes it possible for the dispatcher to "explain" what would otherwise appear to be non-economic dispatches, in ways which are commercially satisfactory to the parties involved.

Thus there is a one-to-one relationship between the binding dual variables and primal constraints. For an AC power system, the simplest OPF problem would be one where the only constraints (other than bounds on generation, voltage, and demand) were the constraints requiring the conservation of active and reactive power. The corresponding dual problem defines the prices of active and reactive power at a node[51] in terms of their marginal system costs with the marginal losses adjusting these prices to produce nodal prices for the independent primal active and reactive power and voltage variables at all other nodes. This is the so called "loss differential solution".   Beyond that, for each binding price constraint added to the model there must be at least one primal constraint added so that the shadow price on that constraint can modify the price pattern so as to explain the observed out of merit order dispatch (OOMOD). In practice, however, if the primal constraint added is not the constraint which actually caused the OOMOD potentially bizarre prices can result. Obviously, the incentives which a dispatcher might have to generate such "explanations", particularly where dispatch was actually sub-optimal, must be carefully considered and controlled. Capacity rights are an appropriate mechanism to provide this discipline, as discussed in Chapter 7.

4.4.4. Relationship Between Independent and Dependent Variables

An examination of the constraints in (4.61) to (4.63) reveals that the price for each of the independent primal variables can be expressed as a linear combination of the prices for the dependent primal variables weighted by the impact which varying the primal independent variable has on each of these primal dependent variables. The prices for the

---

assumptions have not been verified. The issue can be avoided by assuming that reactive power prices are zero at all *PV* nodes.

[51]   Strictly, these prices do not need be defined at the same node.

primal dependent variables are free to vary as required to meet the bounds placed directly on them, or indirectly on them via the bounds on the independent primal variable prices. Thus the situation is, as we might expect, exactly dual to that in the primal problem.

The <u>independent</u> variables for the dual problem are the prices for the <u>dependent</u> variables in the primal problem.

The <u>dependent</u> variables for the dual problem are the prices for the <u>independent</u> variables in the primal problem.

Equations (4.64) to (4.66), which describe the bounds for all the prices, reveal a further symmetry between the primal and dual problem. If a primal variable is not constrained by either of its bounds at optimality, then the dual price associated with its value will be constrained by the fixed, pre-determined marginal costs associated with that primal variable. Indeed, in the absence of cost discontinuities, the price would be constrained to equal a fixed value. Conversely, if the dual price is free to vary outside the range of the marginal costs then one of the multipliers on the primal variable's bounds must be active, implying that the primal variable must be constrained. In the absence of discontinuous costs we could therefore conclude that if a primal variable is constrained then its dual price is unconstrained and vice versa. Cost discontinuities allow prices to take any value between the costs of decreasing and increasing the primal variable. While this may appear to prevent the generalisation of this symmetry it must be remembered that we can divide the range of the primal variable into piece-wise linear sections, each with a fixed marginal cost, producing a problem for which the general result will hold. The relationships between the primal and dual problem are summarised by Table 4.1.

Dispatch Problem: Variables		Pricing Problem: Prices associated with dispatch problem variables.	
Independent	Constrained (or not optimised)	Dependent	Unconstrained
Independent	Unconstrained (and optimised)	Dependent	Constrained
Dependent	Constrained	Independent	Unconstrained
Dependent	Unconstrained	Independent	Constrained

Table 4.1: Relationships between Primal and Dual Variables.

To illustrate how Table 4.1 should be interpreted, consider the case of an independent primal variable, optimised by the dispatcher, which has a capacity limit imposed upon it. If at its observed value the variable is unconstrained (i.e., not up against its capacity limit) then the price associated with that variable must equal the marginal cost of that variables usage[52], and hence the price is constrained. However, the price is functionally dependent on other prices, reflecting the impact of changes in the dispatch variable on other constraints in the network. By contrast, if the capacity constraint on the primal variable is binding, then the price for that variable comprises the (constant) marginal cost plus the capacity rent. While this is also a dependent dual variable, it is no longer constrained in value, the capacity rent taking whatever value is required to maintain consistency with the constrained prices. This implies that an independent primal variable which has not been optimised could be treated as being constrained at its observed solution, the unconstrained shadow price for this constraint providing the pricing model with a mechanism to explain the observed value of the primal variable.

4.4.5. The Impact on Prices of the Scope of the Dispatch

While Table 4.1 classifies the primal OPF problem variables as being either constrained or unconstrained, these classifications depend upon the time frame under consideration. In fact, in the extreme short run all OPF variables can be though of as fixed in value, with the corresponding prices being free to take values which minimise the pricing objective function. At the other extreme, in the very long run, all dispatch variables are essentially unconstrained, and their prices are consequently set to equal some form of "long run marginal cost". In other words, the scope of the pricing regime is fundamentally related to the scope of the dispatch optimisation. This observation has a significant impact on how we explain a dispatch, and is the justification for our claim in the previous chapter that criticisms of the snapshot pricing approach may be mis-directed.

The snapshot model assumed here only accounts for constraints which can be expressed in terms of the variables for that dispatch period, and represents variables which would be unconstrained in a more dynamic representation as being fixed, implying that the associated prices are unconstrained. For instance, a generating unit may be operated below its full capacity so that it can react to support the frequency as load fluctuates. For a snapshot view, the only way to explain the actual level at that instant would be to conclude that its generation level has been externally fixed at the observed level, and the price associated with this "constraint" would thus take whatever value is required to ensure that

---

[52]   Ignoring the issue of discontinuities.

a set of economically consistent prices can be determined. But the loading of this generating unit may be optimal and unbounded in the broader view, and may be explained by a price for "frequency keeping" in that context.

4.4.6. The Prices Actually Charged To The Market

With prices being determined for both independent and dependent primal variables, and for constraints within the model, a further issue arises as to what commodities are actually being bought and sold in the market. Strictly speaking, the optimisation assumes that the dispatcher controls all of the optimisable variables, buying them in at the costs specified by the dispatch problem objective function, and setting them so as to minimise the cost of satisfying the capacity constraints. It therefore seems appropriate to charge (or pay) market participants according to the capacity constraints which are set. For example, the owners of an unconstrained partially loaded generating unit will receive no payment for its capacity, as it is not fully used, and will only be compensated for its fuel costs (priced at marginal cost). The owners of a fully loaded generating unit will receive an additional payment to reward them for making its capacity available. In practice these prices can be combined to form a single "energy price" applicable to both generation and load at that node.

A similar situation occurs with resources made available to the dispatcher by the grid owner. If, as in New Zealand, the grid owner also dispatches the system, then these prices appear as "internal" transfer prices. The total rents received by the dispatcher equal the difference between the total payment made by consumers and the total payment made to generating companies. These rents reflect the marginal costs incurred due to losses, voltage and reactive power constraints, and any general constraints which relate to resources under the control of the grid owner (e.g. transmission capacity). As discussed above, the arbitrary objective function of (4.15) can be used to control these rents.

4.4.7. Impact of Binding General Constraints

If a general constraint involving a new dependent variable is introduced in to this problem then analogous results arise provided that the general constraint is a function of nodal net power injections. The results of Table 4.1 may be obscured, though, in those situations in which a general constraint, being a function of more than one independent variable, becomes binding. The effect of such a constraint is to make one of the independent primal variables into a dependent variable, the general constraint defining its dependence on the remaining independent variables. Consequently, we can use the dual equation corresponding to that variable to substitute for the general constraint multiplier.

In this way the price for this new dependent variable appears with the other dependent variables on the left of (4.61) to (4.63).

4.5. The Pricing Model Developed for New Zealand

Trans Power (NZ) Limited have developed a pricing model based on the approach described here (Read, 1992, Read and Ring, 1995b). Here we report on some of the features of that implementation[53].

In New Zealand, most generating units which are generating active power are simultaneously able to provide (positive or negative) reactive power over a fairly wide range without incurring any additional costs. The reactive power price must therefore be zero at such nodes, if the dispatcher is assumed to have optimised reactive power dispatch, and unless the reactive injection is hard up against a constraint. Thus the New Zealand models assumes that reactive power prices at *PVS* nodes, i.e. all independent reactive power prices, are zero, dropping the corresponding terms out of the expression for dependent prices. Where a reactive constraint is binding at a *PV* node, it is treated as a *PQ* node with a voltage constraint, as in Hogan's model. Reactive power prices at *PQ* nodes, including all non-generator nodes, are unconstrained, since net reactive power injections (loads) there are assumed to have been set externally to the OPF, and not optimised by the dispatcher.

Similarly, voltages at generator nodes are not strictly optimised within the half-hourly time frame, but set according to "voltage profiles" established from experience and off-line studies. To account for this it is currently assumed that all voltage magnitudes are set externally at *PVS* nodes, allowing the model to determine prices for those as necessary. If this were not done, and it was instead assumed that voltage had been optimised, then the requirement that prices for voltage at these nodes be zero would almost certainly preclude any possibility of determining a feasible set of prices. At *PQ* nodes, where voltage magnitude is a dependent variable, the voltage price must be zero if voltage is not constrained, and this is modelled by dropping the relevant terms from all pricing equations.

The objective function used for the pricing model minimises the rents collected by the transmission network, though as noted above, this is ultimately an arbitrary choice.

---

[53]   A recent and updated version of the implemented model described in this section has the ability to relax most of the simplifying assumptions we describe here. At the time of writing, the performance of this new model has yet to be tested.

4.6. Conclusions

In this chapter we have presented an *ex post* approach for determining Dispatch Based Prices for an AC power system dispatch. While based on the approach of Hogan (1991) and Hogan et al. (1995) we consider a more general "PVQ" representation of the electrical relationships between the primal and dual dispatch variables, a step which produces different, and more general, pricing relationships than the "PQ" representation of the earlier works.

Following Hogan we form the dual from the linearisation about an observed OPF solution. Making the assumption that the observed solution is optimal allows us to disregard the "natural" dual objective function in favour of any arbitrary objective function, provided that the complementary slackness conditions are imposed directly on to the dual constraints by removing all terms relating to non-binding primal constraints. We note, though, that some sub-optimally dispatched resources can be treated as if they are at their optimal levels by modelling them as being constrained. We consider this to be a practical approach for addressing sub-optimality in ancillary services not actually traded in the market.

We have analysed the economic interpretation of the prices produced by our model and have shown that they provide incentives consistent with efficient market operation, at least to a "first order" approximation.

In the previous chapter we argued that it was important to carefully define the scope of the responsibilities of the dispatcher. In this chapter we have reinforced that argument by showing mathematically that:

- prices reflect the capacity availability, and hence the owners of capacity must be clearly identified,

- whether a primal variable is considered to be constrained, and hence whether a price must be activated in the dual, depends on the dispatch time frame under consideration and, by implication, the degree of foresight attributed to the dispatcher, and

- whether or not a primal variable can reasonably be considered to be optimised or not depends on whether or not it is considered to be "owned" by the dispatcher.

We have briefly discussed the implementation of the type of model we propose in the New Zealand context, and have illustrated how some of the ideas we describe translate into practice.

Chapter 5

Transmission Constraints[54]

5.1. Introduction

One of the most significant, and most studied, constraints on power system dispatch are transmission constraints. Schweppe et al. (1988) and Hogan (1991, 1992) have explored the basic economic impact of transmission constraints. In this chapter, however, we explore some of the effects arising from more complex situations. In particular, the theory on pricing in transmission systems has often been presented in the terms of acyclic transmission lines, i.e. a single transmission line with a distinct starting point and ending point. For acyclic lines subject to transmission constraints conventional theory suggests a simple economic interpretation, namely that the binding constraint prevents a single marginal generating unit supplying extra power to the regions on the opposite side of the constraint from it, and hence there must be another marginal generating unit on that other side. The node with the most expensive marginal generating unit attached will have a price not less, and typically much greater, than that which would occur if it were possible to send power from the cheaper marginal generating unit. Consequently the transmission constraint raises the cost of operating the system and the marginal impact of this is reflected by an increase in price across the constraint.

The equivalent situation for cyclic transmission lines, i.e. lines which form a loop, is not so easily explained. While we should still expect a price rise across the constraint we must take account of the fact that the "far" side of the constraint is connected with the "near" side by two paths, one directly constrained, the other not[55]. To price in this situation we must introduce the "spring washer" effect.

---

54   This chapter is based on Read and Ring (1995b) and Read and Ring (1995c).

55   Though as we show the binding constraint implicitly effects both paths.

5.2.   Pricing Equations for Transmission Constraints

5.2.1.   The Primal Constraints

Transmission line flows, and the constraints on those flows, can be represented with the general constraint of Chapter 4. Flows can be defined in terms of the independent variables as shown in (5.1) and (5.2).

$$-\overline{P}_k^A\left(\boldsymbol{P}_G^{PX}-\boldsymbol{P}_D^{PX},\boldsymbol{Q}_G^{PQ}-\boldsymbol{Q}_D^{PQ},\boldsymbol{V}^{PVS}\right)+\overline{P}_k \;=0 \qquad\qquad :\eta_{\overline{P}k} \qquad \forall k\in K \qquad (5.1)$$

$$-\overline{Q}_k^A\left(\boldsymbol{P}_G^{PX}-\boldsymbol{P}_D^{PX},\boldsymbol{Q}_G^{PQ}-\boldsymbol{Q}_D^{PQ},\boldsymbol{V}^{PVS}\right)+\overline{Q}_k \;=0 \qquad\qquad :\eta_{\overline{Q}k} \qquad \forall k\in K \qquad (5.2)$$

Here $\overline{P}_k$ and $\overline{Q}_k$ are, respectively, the active and reactive power line flows on transmission line k, where K describes the set of transmission lines. The shadow prices for these constraints are $\eta_{\overline{P}k}$ and $\eta_{\overline{Q}k}$, respectively, and both of these are unconstrained in sign. The bounds on the line flows can be defined in many ways, though we consider two. The first type are simple minimum and maximum megawatt flow limits applied separately to active and reactive power. These are shown in (5.3) to (5.6) with the dual multipliers, which are all non-negative, shown on the right.

$$-\overline{P}_k \;\geq\; -\overline{P}_k^{max} \qquad\qquad :v_{\overline{P}k}^+ \qquad \forall k\in K \qquad (5.3)$$

$$\overline{P}_k \;\geq\; \overline{P}_k^{min} \qquad\qquad :v_{\overline{P}k}^- \qquad \forall k\in K \qquad (5.4)$$

$$-\overline{Q}_k \;\geq\; -\overline{Q}_k^{max} \qquad\qquad :v_{\overline{Q}k}^+ \qquad \forall k\in K \qquad (5.5)$$

$$\overline{Q}_k \;\geq\; \overline{Q}_k^{min} \qquad\qquad :v_{\overline{Q}k}^- \qquad \forall k\in K \qquad (5.6)$$

Note that an upper bound on flow means an upper bound in the conventional direction for that line[56]. An <u>upper</u> bound on flow in the reverse direction is represented by a (negative) <u>lower</u> bound on flow in the conventional direction. Genuine lower bounds on the <u>magnitude</u> of line flows, in each direction, can be represented by positive lower bounds, or negative upper bounds, respectively.

A more physically accurate constraint would be one which jointly constrains active and reactive power flows by limiting the total power flow magnitude, and hence the total heating, on a line. This "thermal" constraint is represented by (5.7) (which is in canonical

---

[56]   That is from node $i(k)$ to node $j(k)$.

form) where T_k^{max} is the square of the maximum power magnitude allowed to flow on line k and χ_k , the shadow price, is non-negative[57].

$$-\overline{P}_k{}^2 - \overline{Q}_k{}^2 \geq -T_k^{max} \qquad\qquad :\chi_k \qquad \forall k \in K \qquad (5.7)$$

This constraint corresponds to requiring that the total MVA (active and reactive) flow on the line must be less than some critical level, T_k^{max}, which may be thought of as the square of the maximum possible complex flow[58].

In practice either a thermal limit or a flow limit would be applied to a single line.

5.2.2. The Dual Pricing Relationships

The general model of (4.15) to (4.24) is shown, with transmission constraints replacing general constraints, in (5.8) to (5.17).   The development of this result is summarised as follows:

- Equation (5.1) to (5.7) are re-arranged into the form of the general primal constraint (4.6), with all terms on the left, and with slack variables inserted to produce an equality constraint[59]. So, for example, (5.7) is re-arranged as:

$$H_k^A\left(\overline{P}_k, \overline{Q}_k\right) = -\overline{P}_k{}^2 - \overline{Q}_k{}^2 + T_k^{max} + \text{slack} = 0$$

- These constraints are substituted into (4.15) to (4.24). For ease of discussion, both types of flow constraints are assumed to exist for each line, though only one flow constraint can be binding.

- Transmission line derivatives with respect to generation and demand are identical, though with opposite sign, hence these derivatives cancel from the price bound constraints, while the derivatives with respect to demand in the equations for the dependent prices can be described as being with respect to net power injections (with a sign change). The form of (4.24) reduces to that of (4.40), and gives rise to (5.17) and (5.18).

---

[57]   This type of constraint is discussed by Hogan (1991), Murphy, Kaye, and Wu (1994), and Read and Ring (1995b).

[58]   Alternatively, this constraint could be scaled so as to represent losses on the left, and a loss limit on the right.

[59]   These slack variables have no impact on the dual problem given that we assume that the binding primal constraints are specified by the observed dispatch.

$$\underset{\substack{\square_P^{\pm K}, \square_Q^{\pm K}, \square^{+K}, \square_P^{+PXS}, \square_Q^{+PXS}, \square_V^{+PXS} \geq \mathbf{0} \\ \square_P^{-PXS}, \square_Q^{-PXS}, \square_V^{-PXS} \geq \mathbf{0} \\ \lambda_P, \lambda_Q, \square_P^K, \square_Q^K, \square_Q^{PV}, \square_V^{PQ}, \square_{\mathbf{P}}^{PXS}, \square_{\mathbf{Q}}^{PXS}}}{\textbf{Minimise}} \qquad \text{Arbitrary Objective Function} \qquad\qquad\qquad\qquad (5.8)$$

subject to :

$$\beta_{Pi} = \lambda_P\left(1 - \frac{\partial L_P^A}{\partial^A P_i}\right) - \lambda_Q \frac{\partial L_Q^A}{\partial^A P_i} - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A P_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A P_i} - \sum_{k \in K}\left(\eta_{\overline{P}k} \frac{\partial \overline{P}_k^A}{\partial^A P_i} + \eta_{\overline{Q}k} \frac{\partial \overline{Q}_k^A}{\partial^A P_i}\right) \qquad \forall i \in PXS \quad (5.9)$$

$$\beta_{Qi} = -\lambda_P \frac{\partial L_P^A}{\partial^A Q_i} + \lambda_Q\left(1 - \frac{\partial L_Q^A}{\partial^A Q_i}\right) - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A Q_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A Q_i} - \sum_{k \in K}\left(\eta_{\overline{P}k} \frac{\partial \overline{P}_k^A}{\partial^A Q_i} + \eta_{\overline{Q}k} \frac{\partial \overline{Q}_k^A}{\partial^A Q_i}\right) \qquad \forall i \in PQS \quad (5.10)$$

$$\beta_{Vi} = -\lambda_P \frac{\partial L_P^A}{\partial^A V_i} - \lambda_Q \frac{\partial L_Q^A}{\partial^A V_i} - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A V_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A V_i} - \sum_{k \in K}\left(\eta_{\overline{P}k} \frac{\partial \overline{P}_k^A}{\partial^A V_i} + \eta_{\overline{Q}k} \frac{\partial \overline{Q}_k^A}{\partial^A V_i}\right) \qquad \forall i \in PVS \quad (5.11)$$

$$\mu_{Qi} = \beta_{Qi} - \lambda_Q \qquad\qquad\qquad\qquad \forall i \in PV \quad (5.12)$$

$$\mu_{Vi} = \beta_{Vi} \qquad\qquad\qquad\qquad \forall i \in PQ \quad (5.13)$$

$$c_{Pi}^- - \left\langle \upsilon_{Pi}^- \right\rangle \leq \beta_{Pi} \leq c_{Pi}^+ + \left\langle \upsilon_{Pi}^+ \right\rangle \qquad\qquad\qquad\qquad \forall i \in PXS \quad (5.14)$$

$$c_{Qi}^- - \left\langle \upsilon_{Qi}^- \right\rangle \leq \beta_{Qi} \leq c_{Qi}^+ + \left\langle \upsilon_{Qi}^+ \right\rangle \qquad\qquad\qquad\qquad \forall i \in PXS \quad (5.15)$$

$$\beta_{Vi} = \left\langle \upsilon_{Vi}^+ \right\rangle - \left\langle \upsilon_{Vi}^- \right\rangle \qquad\qquad\qquad\qquad \forall i \in PXS \quad (5.16)$$

$$\eta_{\overline{P}k} = \left\langle \upsilon_{\overline{P}k}^+ \right\rangle - \left\langle \upsilon_{\overline{P}k}^- \right\rangle + 2\overline{P}_k^* \left\langle \chi_k \right\rangle \qquad\qquad\qquad\qquad \forall k \in K \quad (5.17)$$

$$\eta_{\overline{Q}k} = \left\langle \upsilon_{\overline{Q}k}^+ \right\rangle - \left\langle \upsilon_{\overline{Q}k}^- \right\rangle + 2\overline{Q}_k^* \left\langle \chi_k \right\rangle \qquad\qquad\qquad\qquad \forall k \in K \quad (5.18)$$

Equations (5.17) and (5.18) are the dual constraints corresponding to the capacity charges on the primal line flows, $\overline{P}_k$ and $\overline{Q}_k$. We investigate the implications of these to each of the two forms of transmission constraint separately. If we have only flow limit constraints these equations become:

$$\eta_{\overline{P}k} = \left\langle \upsilon_{\overline{P}k}^+ \right\rangle - \left\langle \upsilon_{\overline{P}k}^- \right\rangle \qquad\qquad \forall k \in K \qquad (5.19)$$

$$\eta_{\overline{Q}k} = \left\langle \upsilon_{\overline{Q}k}^+ \right\rangle - \left\langle \upsilon_{\overline{Q}k}^- \right\rangle \qquad\qquad \forall k \in K \qquad (5.20)$$

That is, the value associated with the flow on a line is defined by the dual variables corresponding to the upper and lower bounds on the flow. Complementary slackness requires that if a primal constraint is not binding then the corresponding dual variable is zero. Thus, for active power, $\upsilon_{\overline{P}k}^+$ and $\upsilon_{\overline{P}k}^-$ can not both be non-zero. If active power flow is not at its upper bound then $\upsilon_{\overline{P}k}^+ = 0$, and if it is also not at its lower bound then $\upsilon_{\overline{P}k}^- = 0$ and hence, $\eta_{Pk} = 0$. That is, if the flow on a line is not constrained by a bound, then there is no value in having any more line capacity. However, if the line flow is at an upper bound (from node i to j, i.e. in the conventional direction specified for the line) then $\upsilon_{\overline{P}k}^+ \geq 0$, indicating that, given one <u>more</u> unit of line capacity in that direction, we should be able to <u>reduce</u> the cost of running the system[60], and hence $\eta_{Pk} = \upsilon_{\overline{P}k}^+ \geq 0$, indicating the maximum amount we would be prepared to pay to have one more unit of line flow.

In reaching this conclusion it must be remembered that the upper bound on active power line flows, equation (5.3), is in canonical form[61], with the negative of the line capacity on the right, and since $\upsilon_{\overline{P}k}^+$ equals the change in the cost of running the system resulting from an increase in the right hand side, $\upsilon_{\overline{P}k}^+$ actually measures the marginal value of <u>decreased</u> line capacity. Hence a decrease in the line capacity increases the system operating cost ($\eta_{Pk} = \upsilon_{\overline{P}k}^+ \geq 0$) so increased line capacity must decrease the cost of operating the system. Therefore, η_{Pk} indicates the value of having one more unit of line capacity.

Similarly, if flow is at a lower bound in that direction, then $\eta_{\overline{P}k} = -\upsilon_{\overline{P}k}^- \leq 0$, which indicates that if the lower bound were <u>raised</u> by one unit in that direction then the cost of

---

[60]   Note that these multipliers multiply derivatives with respect to power <u>injection</u>. These terms are then <u>subtracted</u> from the expression for the price for power <u>extracted</u> from the network.

[61]   And remains in canonical form when substituted for the general primal constraint.

running the system would <u>increase</u>, and we would have to be <u>paid</u> before we would want to reduce the line flow[62]. Identical conclusions apply for reactive power line flows.

If we have only thermal limit constraints then (5.17) and (5.18) reduce to:

$$\eta_{Pk} = 2\overline{P}_k^* \langle \chi_k \rangle \qquad\qquad \forall k \in K \qquad (5.21)$$

$$\eta_{Qk} = 2\overline{Q}_k^* \langle \chi_k \rangle \qquad\qquad \forall k \in K \qquad (5.22)$$

If, at optimality, a thermal limit constraint is not binding then $\eta_{\overline{P}k} = \eta_{\overline{Q}k} = \chi_k = 0$, as there is no value in having any more thermal capacity because there is no use for it. However, if the thermal limit becomes binding then $\chi_k > 0$ and $\eta_{\overline{P}k} = 2\overline{P}_k^* \chi_k > 0$ (assuming positive active power line flow in the conventional direction), indicating that a unit increase in our ability to send active power through the line, in that direction, would reduce the cost of operating the system, and hence we would be prepared to pay up to a positive amount, $\eta_{\overline{P}k}$, for that unit. Similarly, if we could send one more unit of reactive power flow through the line, the value of that extra unit would be $\eta_{\overline{Q}k} = 2\overline{Q}_k^* \chi_k > 0$ (again assuming positive line flows). If the flow is at a maximum in the reverse direction then the signs of the capacity prices are reversed, but the interpretation is consistent.

It is worth noting that while the transmission flow constraints stated here encompass four possibilities (upper or lower bounds in each direction), it is not necessary for the user of such a model to specify which is intended. Recall that each line has a specified "conventional direction" of flow, and that we model "reverse" flows as negative flows in the conventional direction. If a constraint is binding it must impose a cost on the system by impeding flow from a lower priced region to a higher priced one. In principle, we could imagine solving the model to produce an unconstrained "loss differential" solution[63], deducing the type of constraint required, then re-solving the model with the constraint imposed. In practice, only one of an upper/lower constraint pair can have a positive multiplier in any one dispatch. (This condition still holds if both upper and lower bound constraints are set to the same level, and hence are both "binding", as only one of the constraints will have a positive multiplier associated with it).

Although such a model must interpret the constraint in absolute terms, the actual observed line flow (from the load flow solution) is known, and thus the model can

---

[62] If the lower bound is really an upper bound in the reverse direction, these terms will have opposite signs and hence <u>raising</u> the bound corresponds to <u>restricting</u> the flow in that direction.

[63] With prices defined relative to only one marginal node while ignoring the price bounds imposed by other marginal nodes.

automatically determine whether a constraint was in the forward or reverse direction. Thus both multipliers can safely be activated in a model, leaving the model to determine which is required to explain the price discrepancy observed in the dispatch[64].

5.3. The Impact on Prices of Transmission Constraints

5.3.1. Introduction

In this section we describe some of the results arising from experimentation with Trans Power's AC pricing model, and present an economic analysis of those results. While it has long been recognised that the impact of transmission constraints is to raise the price of power at the receiving end of a transmission line while lowering it at the sending end, little work has been reported on the impact of such constraints in more complex system of lines, particular cyclic systems. The work we present goes some way towards rectifying this situation.

It should be noted that Wu et al. (1994) and Oren et al. (1995) report a number of interesting pricing phenomenon, including a phenomenon similar to the "spring washer" effect discussed below, resulting from binding constraints in cyclic transmission lines. However, their results are based on the assumption that the complementary slackness conditions applying to the constraints on transmission line flow can be ignored. This assumption is introduced because Wu et al. assume that the transmission system is not a "profit-maximising, competitively organised goods-transporting sector"[65]. A consequence of this assumption is that their is no longer a constraint imposing consistency in nodal prices, and hence their results are not surprising. We make no such assumption in this work, with an optimal, cost minimising dispatch assumed, an assumption stemming from the centralised pool dispatch process that we consider. In fact, the results of Wu et al. would, in our interpretation, correspond to a sub-optimal dispatch.

---

[64]   Strictly this logic only applies with the original dual objective function, and since including both multipliers violates complementary slackness, the correct choice can not be guaranteed with a revised objective function. More complex logic could be devised to overcome this problem, but it is certainly irrelevant in the situation where constraints are only introduced as required to explain price discrepancies. In that case prices can definitely only move in one direction, and only the correct multiplier can achieve this.

[65]   Wu et al. attempt to refute the results of Hogan (1991, 1992) on the (incorrect) premise that it treats electricity networks as if they behave like transportation networks.

5.3.2. Prices Observed in New Zealand With Transmission Constraints

Figure 5.1 depicts a sub-section of the North Island transmission grid. A loop comprising a 220kV system is shown, as well as an acyclic 110kV system running up the east coast of the North Island. Using a sample dispatch for the full North Island system, Trans Power's AC *ex post* pricing model was solved in the absence of transmission constraints, as discussed by Drayton, Read, Ring and Rosevear (1993). The resulting "loss differential" solution explains all price variations in terms of marginal losses. As there were no constraints, only one generating station, Huntly, was marginal having a price set at $30/MW. To simulate the impact of a (simple active power flow limit) transmission constraint, the pricing model was re-solved for the same dispatch, but with the price set to a new, non-optimal value at a node other than Huntly, effectively indicating to the model that this is a second marginal node. The transmission constraint multipliers on one of the transmission lines were activated so as to allow the model a way of explaining this apparent out-of-merit-order dispatch (the marginal transmission flow could be in either direction). The resulting nodal prices were recorded for each of a range of price settings at the second "marginal" node.
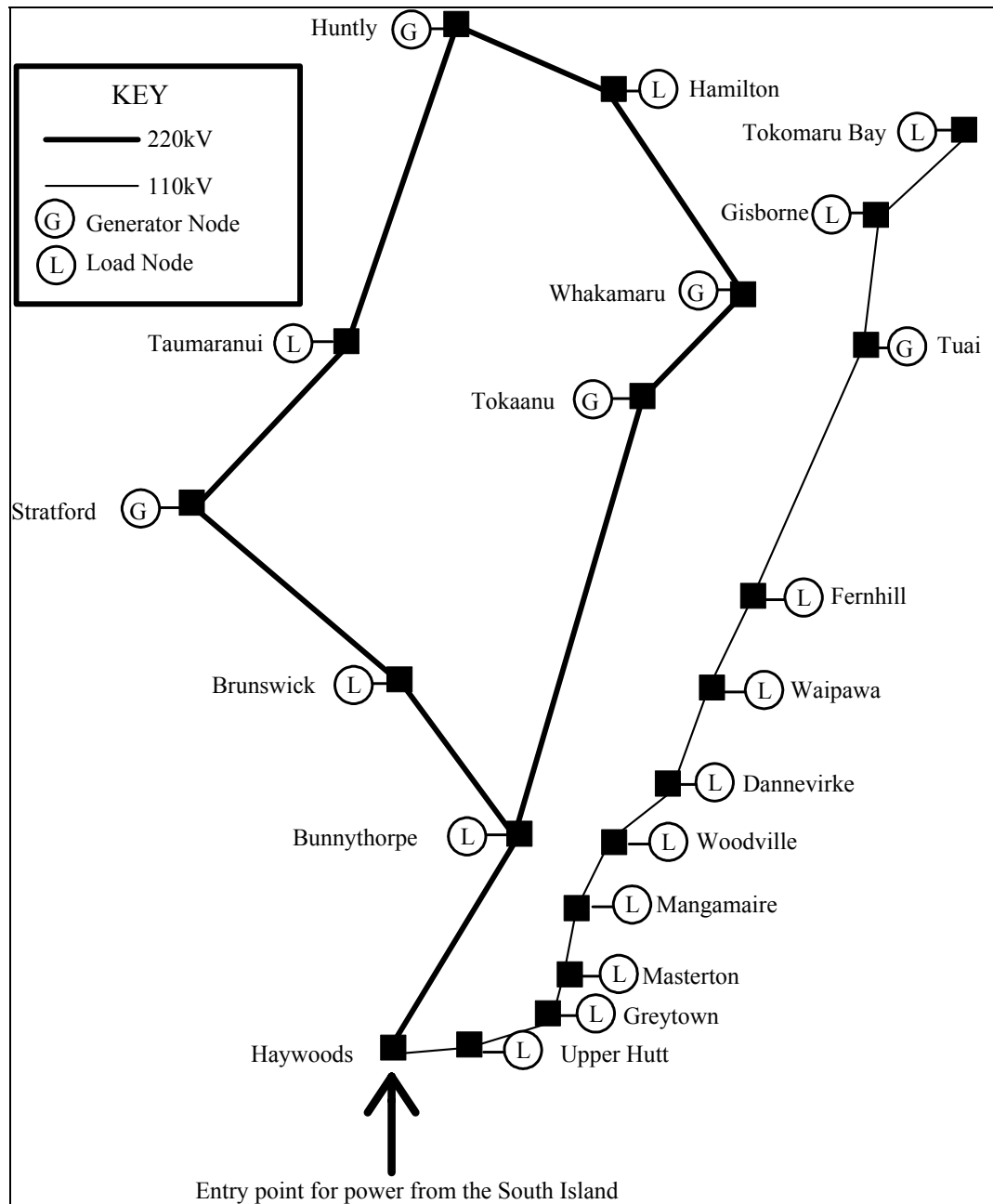
Figure 5.1: A sample of the North Island transmission system

The optimal, loss differential price at Tuai, a node on the 110kV acyclic line, was $28.5/MW. A constraint on the line joining Fernhill and Tuai was simulated by setting the price at Tuai to values ranging between $0/MW and $50/MW and activating the constraint multiplier for the Fernhill-Tuai line. The resulting prices along the acyclic stretch of line are shown in Figure 5.2.
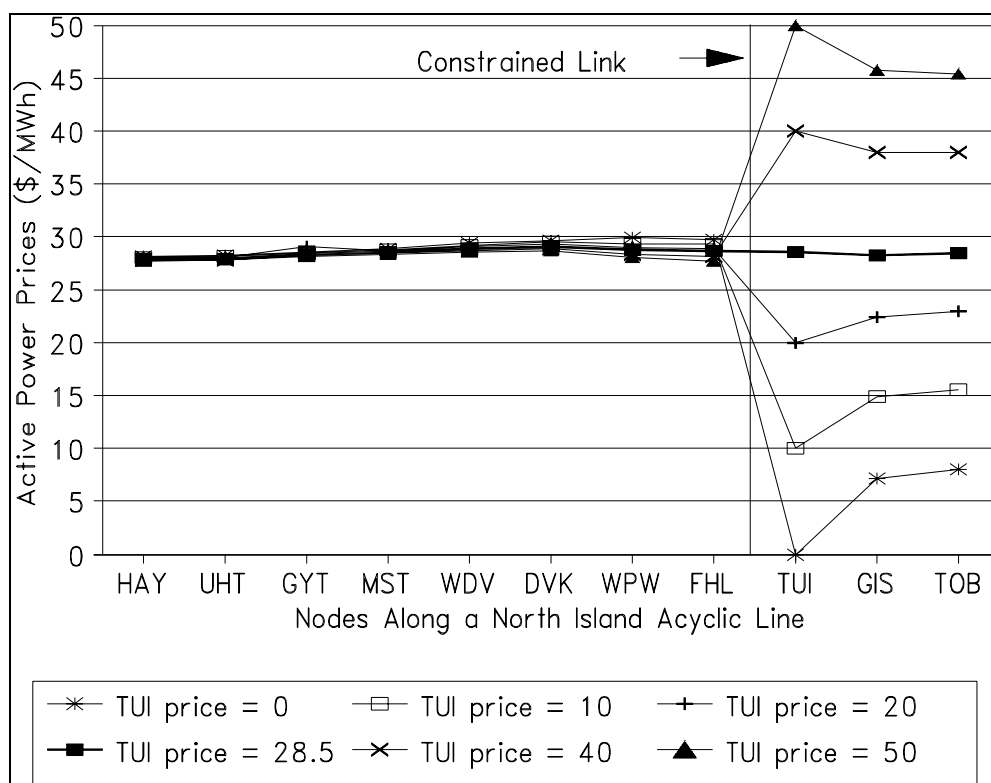
Figure 5.2: Active power prices along an acyclic line in North Island

Figure 5.2 demonstrates the generally reported impact of transmission constraints, that is, that prices fall at the sending end, and rise at the receiving end, of a constrained transmission line[66]. The loss differential solution is indicated by the bold line.

A more interesting situation arises when a constraint is activated within the 220kV loop. The optimal, loss differential price at Tokaanu was $28.3/MW. In a second experiment, the transmission constraint multipliers on the link joining Tokaanu to Whakamaru were activated, with the price at Tokaanu being varied from its optimal value down to $0/MW[67], this out-of-merit-order dispatch helping to relieve this constraint, thus making Tokaanu a second marginal node. Trans Power's pricing model determined the new prices for the full North Island network, the effect of the constraint on the active power prices around the loop being shown in Figure 5.3. The unconstrained loss differential prices are shown by

---

[66]   The marginal flow is from Tuai to Fernhill when the price at Tuai is low, and from Fernhill to Tuai when it is high.

[67]   Price increases at Tokaanu are not considered as the pricing model would be unable to determine prices in this situation. This is because an increase in price at Tokaanu would imply that a marginal unit of power should flow from Huntly towards Tokaanu and this would only worsen the condition of the constrained line. Hence such a dispatch could not be optimal and there is no consistent set of prices that could explain such a dispatch.

the bold line. As the Tokaanu price is lowered, other prices are affected, and either pushed higher or lower than their loss differential price. The same result does not apply, however, for reactive power (the constraint is a simple flow limit which is independent of reactive power), as shown by Figure 5.4[68].
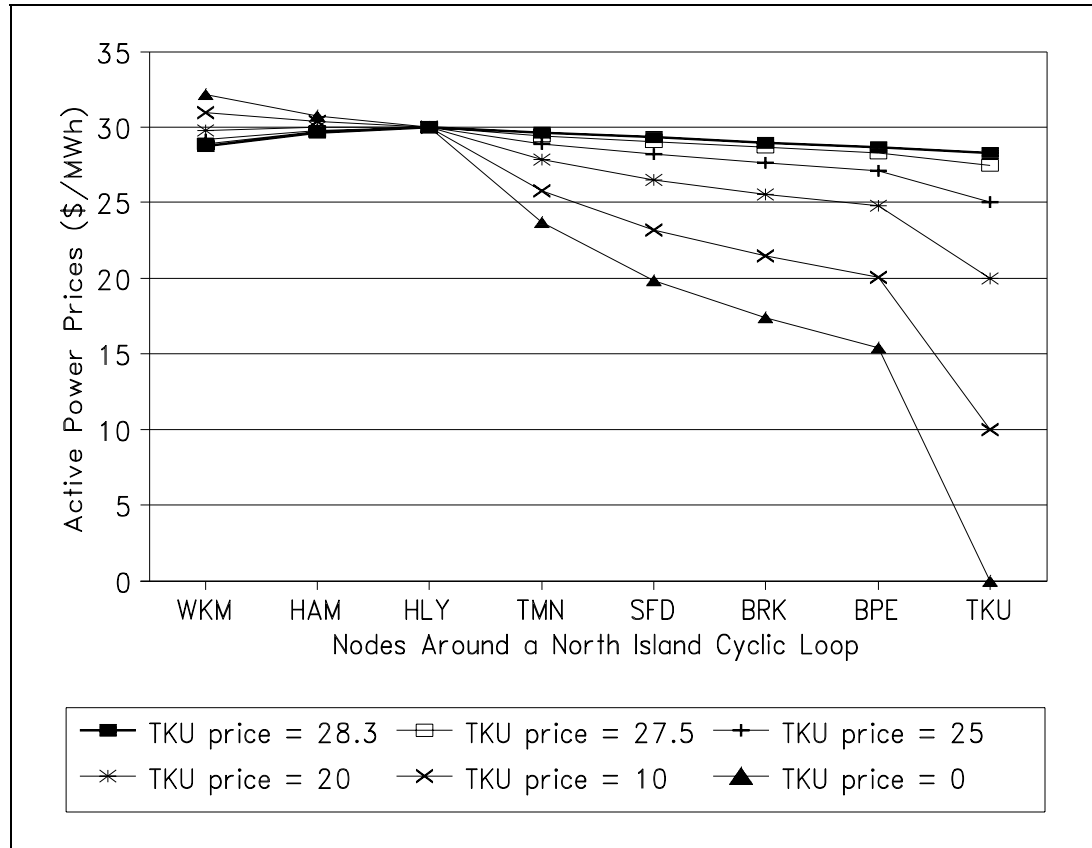
Figure 5.3: Active power prices around a cyclic loop in North Island

---

[68]   We have arbitrarily superimposed a constraint on an optimal unconstrained dispatch, hence the results depicted in Figure 5.4 assume the same reactive power injections as for the unconstrained dispatch. In practice, however, reactive power injections should be re-optimised so as to take account of this constraint Consequently, the prices shown in Figure 5.4. are significantly higher (possibly by several orders of magnitude) than would be the case had reactive power injections been optimised with the transmission constraint imposed. Kahn and Baldick (1994) use similar examples to highlight the importance of ensuring that the dispatch upon which prices are based is in fact optimal.
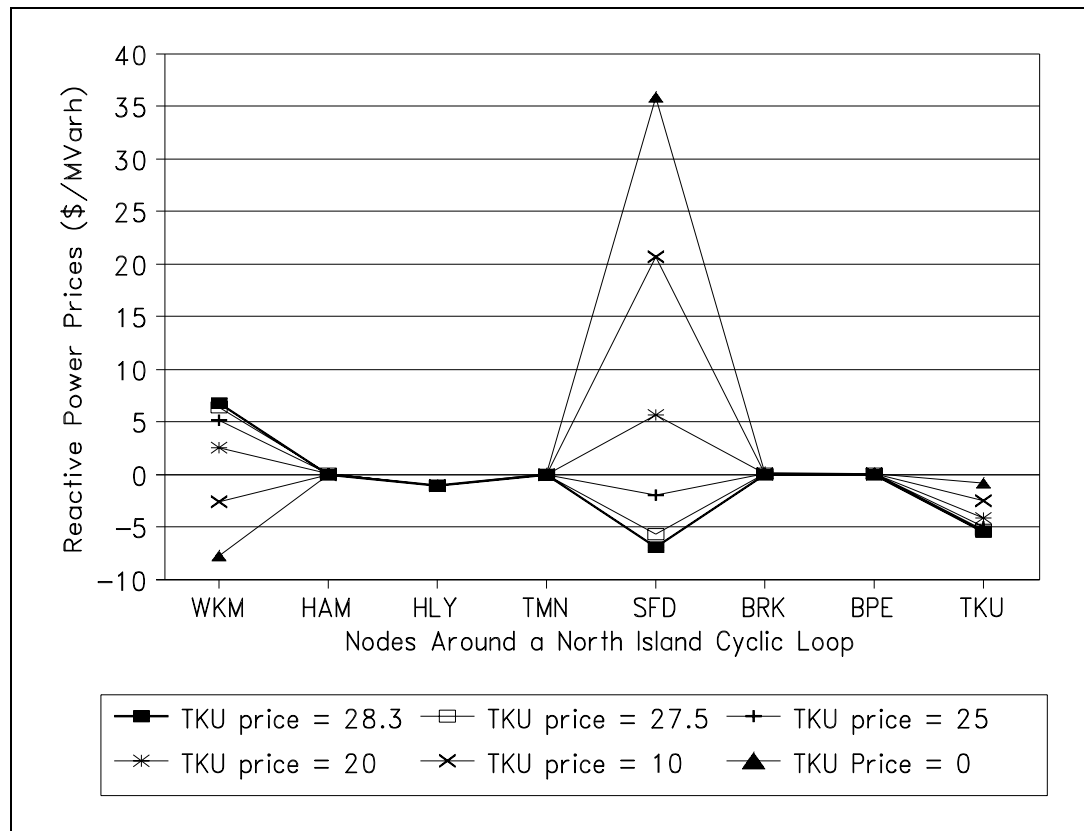
Figure 5.4: Reactive power prices around a cyclic loop in North Island

It is apparent from Figure 5.3 that the price of power rises around the network as we move from the Tokaanu side of the constraint around towards Huntly and on to Whakamaru. This implies a significant drop in price as we continue across the constrained line from Whakamaru back to Tokaanu. We refer to this phenomenon as the "spring washer" effect. For acyclic lines prices at nodes receiving power through a constraint only ever rise compared with unconstrained loss differential prices. This spring washer effect demonstrates that in constrained acyclic loops prices can either rise or fall compared to their loss differential values.

5.3.3. Analysis of the Spring Washer Effect

To explain the spring washer effect we must first explore the behaviour of transmission prices about an unconstrained cyclic transmission loop. Figure 5.5 depicts a simple transmission loop with power flowing in one side and out the other side.
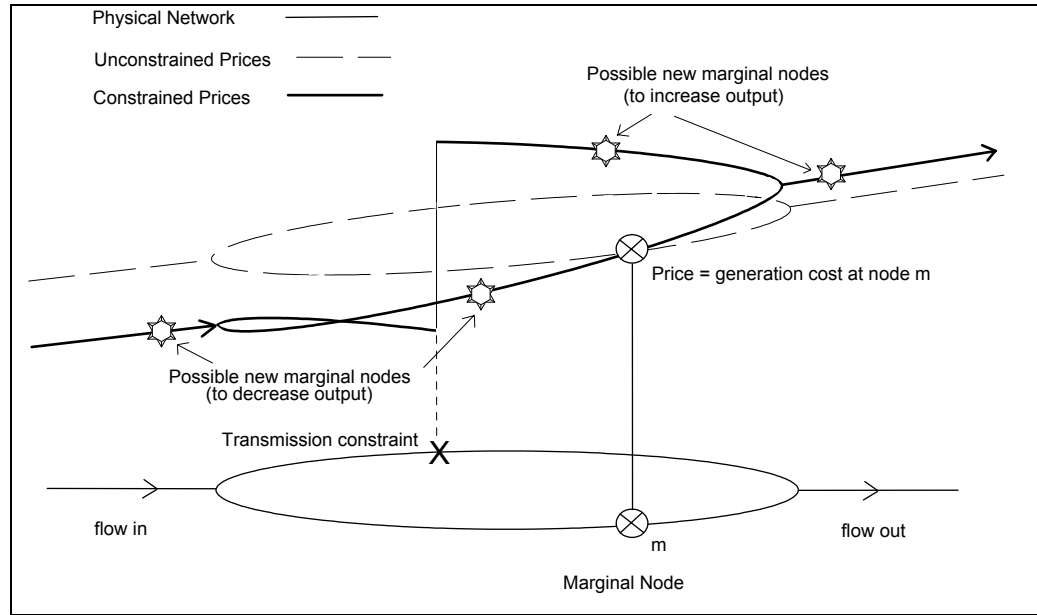
Figure 5.5: The "Spring Washer Effect"

The marginal source of power is at node *m*. There could either be a partially loaded generating unit at this node or a transmission line which connects node *m* with a partially loaded generating unit outside of the loop. Current entering the loop has a choice of two paths to take around the loop and so splits itself between the paths, as required by Kirchhoff's Laws of current and voltage, so as to follow the path of least impedance. If we assume that power flows on each path are both related to their currents by the same constant[69], and since current is inversely proportional to impedance, the ratio of the power flows travelling each path must conform to the relationship (Read and Ring, 1995c):

$$\left(\frac{\text{power flow on path 1}}{\text{power flow on path 2}}\right) = \left(\frac{\text{impedance on path 2}}{\text{impedance on path 1}}\right) \tag{5.23}$$

For the unconstrained problem the relative prices around the loop only depend on losses. The prices, indicated by the closed line, rise across the loop to reflect the losses incurred in getting power from the less expensive source on the left side of the loop to the more expensive right side of the loop. As more power is required on the right side of the loop, more power is injected at the marginal node with the flows of that power dividing according to the equation above.

---

69   This is only really true if the only two voltages on the loop are at the points where flows enter and exit the loop, or if the voltage profile around the loop is consistent with there being only two voltages set.

Imposing a transmission constraint on the loop immediately implies that flow cannot be increased across the constraint, and that it is impossible for one marginal generating unit (or generating station) alone to provide extra power without increasing the flow on the constrained line. Thus one new marginal generating unit is required. The solid line in Figure 5.5 depicts the situation which is now encountered. The changes can be summarised as follow:

- The price jumps at the constrained point, simultaneously falling on the upstream side and rising on the downstream side.

- The price remains constant at the point where power injected from the current marginal node (m) enters the loop.

- Somewhere around the loop, or connected to the loop, a new marginal node appears.

- If it is connected between the current marginal node and the upstream side of the constraint, the new marginal node will have to back off its net injection into the loop so as to relieve pressure on the upstream side of the constraint.

- If it is connected between the current marginal node and the downstream side of the constraint, the new marginal node could will now have to inject more power into the loop to relieve pressure on the constraint by increasing flow to its downstream side, and so backing off flow across it.

- The price at the point where power injected from the new marginal node enters the loop must reflect its marginal cost.

- From the original marginal node, where the price remains constant, the price rises smoothly around the loop, to the downstream side of the constraint, and falls in the other direction, to the upstream side of the constraint, where it can even become negative[70].

The pricing resulted depicted in Figure 5.5 assumes that power flows in at one point in the loop and out at another, without any other in-takes or off-takes. Consequently, losses rise continuously from the injection point around to the off-take point, giving rise to the smoothness of the price curves in Figure 5.5. In practice, however, there may be other points at which power enters or leaves the loop, with power flows travelling in different

---

[70]   This reflects a situation in which the dispatcher would be prepared to pay for increased load as the best way to relieve pressure on a link, a situation which is not physically impossible.

directions on different segments of the loop. As a consequence the smooth price profile may be replaced by an undulating profile, with the effect on prices of losses and the transmission constraint reinforcing one-an-other at some points, while countering each other at other points. However, regardless of where injections occur, the impact on the loop prices due to the transmission constraint will be reflected throughout the network, depending on where each branch is ultimately connected into the loop. Where branches are connected in such a way as to form parallel loops, prices will rise around those parallel loops in a similar manner.

A derivation of prices around a transmission loop, assuming a simple DC load flow, is presented by Read and Ring (1995c). Read and Ring consider a loop like that in Figure 5.6, with nodal power injections indicated by P_i and resistances (the DC form of impedance) between nodes i and j defined by R_{ij}. The "observed" solution was defined to be one with all voltages being identical (i.e. there were no flows), with prices being derived by perturbing the nodal voltages individually.
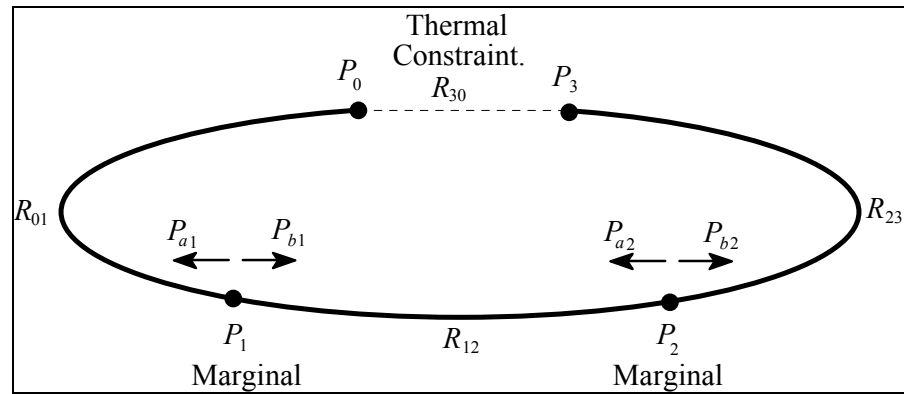
Figure 5.6: A constrained transmission loop with nodal "position" defined by cumulative resistance.

This diagram represents a DC approximation where:

- Line resistance between nodes i and j is indicated by R_{ij} while the nodal net injection at node i is indicated by P_i.

- The line connecting nodes 0 and 3 is subject to a thermal constraint. This constraint effectively places a limit on the current that can flow through the line.

- When the thermal constraint is not binding the generating unit at node 2 is the only marginal source of power. When the constraint becomes binding a second marginal generating unit is required, this is taken to be the generating unit at node 1.

- For simplicity, we ignore losses, so the flows out of the injection nodes equal the flows into the consumption nodes.

The resulting power price at some point X is:

$$\beta_{PX} = \left(\frac{\beta_{P1} + \beta_{P2}}{2}\right) - \left(\beta_{P1} - \beta_{P2}\right)\left(\frac{R_X}{R_{12}}\right) \qquad (5.24)$$

Here R_X measures the "electrical distance" (i.e. the resistance) between the point X and the point electrically half way between the marginal generators, i.e. at an electrical distance of $R_{12}/2$ from each[71]. R_X has a negative value at node 0, and a positive value at node 3. Table 5.1 illustrates the application of (5.24) to the nodes in Figure 5.6. Whether the prices increase or decrease from one side of the constraint to the other depends solely on the sign of the difference between the prices at the two marginal nodes.

Node	Position: R_X	Price: $\beta_P(R_X)$
0	$-\dfrac{R_{12}}{2} - R_{01}$	$\beta_{P0} = \beta_{P1} - \left(\beta_{P2} - \beta_{P1}\right)\left(\dfrac{R_{01}}{R_{12}}\right)$
1	$-\dfrac{R_{12}}{2}$	β_{P1}
2	$\dfrac{R_{12}}{2}$	β_{P2}
3	$\dfrac{R_{12}}{2} + R_{23}$	$\beta_{P3} = \beta_{P2} + \left(\beta_{P2} - \beta_{P1}\right)\left(\dfrac{R_{23}}{R_{12}}\right)$

Table 5.1: Prices around a constrained cyclic loop.

The constraint shadow price is given by the equation:

$$\eta_{\bar{P}30} = \left(\beta_{P1} - \beta_{P2}\right)\left(\frac{R_T}{R_{12}}\right) \qquad (5.25)$$

Here R_T is the total resistance around the loop, that is:

---

[71] It is not strictly necessary to use the midway point" as a reference and the formula can be simplified if we use one of the marginal nodes instead. However, the choice of reference is arbitrary, and the form used here highlights the symmetry of the relationship.

$$R_T = R_{01} + R_{12} + R_{23} + R_{30} \qquad (5.26)$$

The following observations can be made about these results:

- The extent of the spring washer price effect is ultimately proportional to:

  - the difference between the marginal costs at the two marginal nodes,

  - the electrical "distance" around the loop (i.e. total loop impedance), and

  - the <u>inverse</u> of the electrical "distance" between the two marginal nodes via the unconstrained path[72].

- If, for example, $\beta_2 > \beta_1$ (and none of the line resistances are zero) it is apparent that $\beta_3 > \beta_2 > \beta_1 > \beta_0$.

- The absolute position of the marginal nodes on the loop, relative to the constraint, do not influence the value of $\eta_{\bar{P}30}$ if the electrical distance (R_{12}) between them is fixed. The prices will change around the loop, however.

- If the marginal nodes lie outside the loop, these same formulae can be applied to the points at which their output enters the loop[73]. Naturally, costs are adjusted for losses from the point of generation to the point of entry to the loop.

- $\eta_{\bar{P}30}$ is <u>not</u> the price differential across the constraint. Instead it corresponds to the price jump which would occur if we were to imagine the constraint as occurring at some point within the constrained line. We can imagine the price equation as applying at all points along the constrained line, continuing to rise from the downstream side, and fall from the upstream side, until the price jump occurs at the hypothetical constrained point[74]. This situation is shown in Figure 5.7.

---

[72]   Thus it is proportional to the ratio of the resistance between the marginal generators to the total loop resistance. This corresponds to the engineering concept of "leverage". This relates to the effectiveness of dispatching the generators in a manner so as to keep flows across the constraint constant. If the two marginal generators are electrically close to each other, quite large adjustments will have to be made in order to control the distant flow. This makes it difficult, and expensive, to achieve control, and produces correspondingly large, positive and negative, price shifts in our model.

[73]   Multiple loops are discussed below.

[74]   Read and Ring (1995c) give a derivation but the result should be obvious from the fact that, by introducing notional nodes, we can make the "constrained line" appear either longer, or shorter, without affecting the real situation. As the line(s) which are deemed to be constrained extends (contracts) to include more (less) of the loop, the differential across the line falls (rises), according
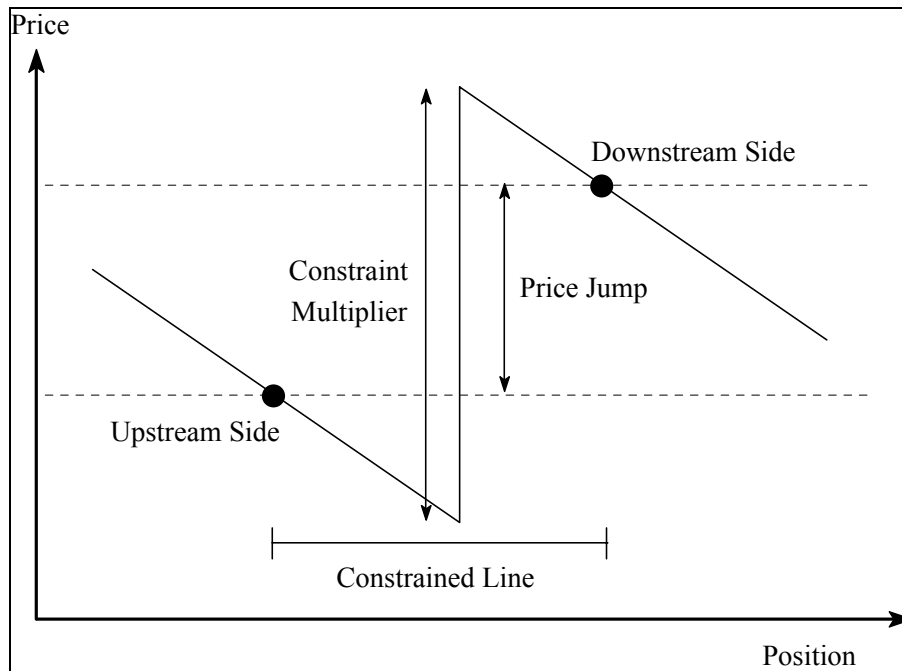
Figure 5.7: The constraint shadow price versus the price jump.

- As a result of these loop effects, active power prices can <u>fall</u> in the direction of active power flow.

- Prices for points on the upstream side of the loop, between the upstream side of the constraint and the cheaper marginal node, will be lower than the marginal cost of power at either marginal node.

- Prices for points on the downstream side of the loop, between the downstream side of the constraint and the more expensive marginal node, will be higher than the marginal cost of power at either marginal node.

- Prices can even become negative on the upstream side at the constraint, implying that we should be prepared to pay generators to reduce output, or consumers to increase load, if that is the best way to reduce pressure on the constrained line.

- Prices for points between the two marginal nodes, on the unconstrained side of the loop will lie between the prices at the marginal nodes.

---

to the formula, with the highest differential, $\eta_{\bar{P}30}$, applying to a line of infinitesimal length, including the constraint.

- Congestion rentals apply between all points in the constrained loop[75], and these may be positive or negative depending on the influence which the indirect "loop flows"[76] occasioned by such transfers have on the constrained line.

- Any transfer between points outside the loop, but which must pass through the loop, will incur positive or negative congestion rentals depending on its point of entry and exit.

5.3.4. A Numerical Example of the Spring Washer Effect

Figure 5.8 depicts a greatly simplified representation of the central North Island loop. We assume that the only impedance is line resistance (represented here, very approximately, in relative terms). This dispatch is unconstrained, but the line joining Bunnythorpe to Huntly has just reached its maximum flow level[77]. Huntly is generating at full capacity but requires an additional 100MW from the south to meet the load in the north. Bunnythorpe, receiving power from the South Island, is marginal with a price of $10/MW. Ignoring losses, this price holds throughout the North Island network. The power injected at Bunnythorpe splits between the two available paths, with 50MW going each way, as dictated by the line resistances. Stratford is not generating because its marginal cost of $20/MW exceeds the price there for this unconstrained dispatch.

---

[75] Although there is only a shadow price on the constrained line itself.

[76] If power flows through one circuit then the physical laws of current flow will require that power flows along attached parallel circuits. These parallel flows are "loop flows".

[77] We here assume that there is no demand for increased line flow, so the fact that the line flow is at its upper limit does not actually constrain the dispatch.
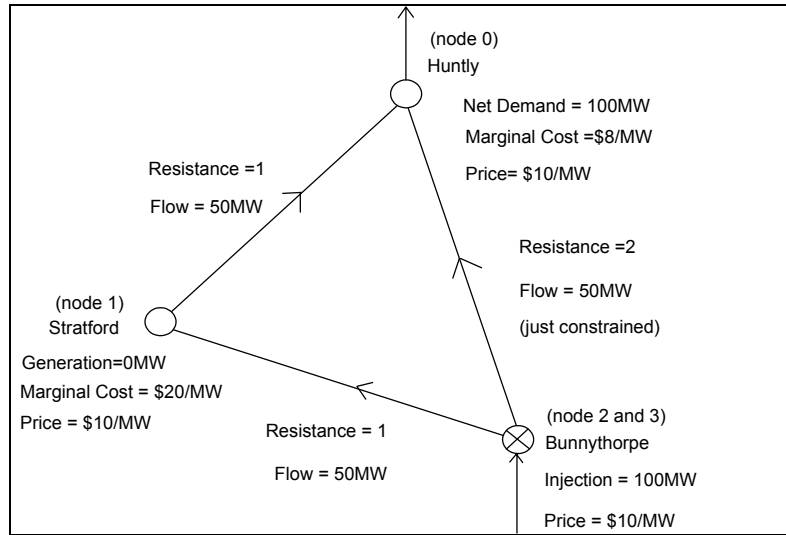
Figure 5.8: An unconstrained three node network.

The line joining Bunnythorpe to Huntly now becomes constrained. We want to determine the new prices at Huntly (which is not marginal) and the value of extra capacity on the constrained line. To get more power to Huntly the generating station at Stratford must play a role, and hence both Stratford and Bunnythorpe must be marginal nodes. The fact that power is flowing through a constrained loop, towards Huntly immediately indicates that Huntly must have a price higher than at either of the other two nodes.

In terms of Figure 5.6, Huntly corresponds to node 0, Stratford to node 1, and Bunnythorpe to node 2. Nodes 2 and 3 are at the same point in space as there is no line between them. Hence we have:

$$\beta_{P1} = \$20/\text{MW (Stratford)} \qquad\qquad \beta_{P2} = \$10/\text{MW (Bunnythorpe)}$$

$$R_{01} = 1 \text{ (Huntly-Stratford)} \qquad\qquad R_{12} = 1 \text{ (Stratford-Bunnythorpe)}$$

$$R_{23} = 0 \text{ (No line)} \qquad\qquad R_{30} = 2 \text{ (Bunnythorpe-Huntly)}$$

$$R_T = 1 + 1 + 2 = 4$$

R_X is zero at a point half way between Stratford and Bunnythorpe and decreases as we move towards Stratford. Hence, at Huntly, we have:

$$R_X = -\frac{R_{12}}{2} - R_{01} = -\frac{1}{2} - 1 = -1.5$$

Using the spring washer effect equation we have a price at Huntly of:

$$\beta_{PX} = \left(\frac{\beta_{P1} + \beta_{P2}}{2}\right) - (\beta_{P1} - \beta_{P2})\left(\frac{R_X}{R_{12}}\right)$$

$$= \left(\frac{20+10}{2}\right) - (20-10)\left(\frac{-1.5}{1}\right)$$

$$= \$30/MW$$

While, the value of extra line capacity is:

$$\eta_{\bar{P}30} = (\beta_{P1} - \beta_{P2})\left(\frac{R_T}{R_{12}}\right)$$

$$= (20-10)\left(\frac{4}{1}\right)$$

$$= \$40/MW$$

That is, we would be prepared to pay \$40 if we were to be able to send one more megawatt of power along the line.

These results can be derived more intuitively, however. Let us suppose that demand at Huntly has risen by 10MW. The generating station at Huntly is fully loaded, and cannot meet this extra demand. Bunnythorpe cannot supply this extra load on its own, as this will cause the constraint to be violated. Hence Stratford must increase generation to supply this demand, with the power flows split across the two paths being determined by (5.23):

$$\left(\frac{\text{power flow}_{\text{Stratford-Bunnythorpe-Huntly}}}{\text{power flow}_{\text{Stratford-Huntly}}}\right) = \left(\frac{\text{resistance}_{\text{Stratford-Huntly}}}{\text{resistance}_{\text{Stratford-Bunnythorpe-Huntly}}}\right)$$

$$= \left(\frac{1}{1+2}\right) = 0.333$$

That is, the flow from Stratford to Huntly via Bunnythorpe is just one third of the amount that flows directly to Huntly. Hence for every additional megawatt injected from Stratford the flow across the constrained link would be increased by 0.25MW, while the flow directly to Huntly would increase by three times this, or by 0.75MW.

Because of the line constraint, Bunnythorpe must decrease its injection so as to allow the extra 0.25MW flow produced by each additional unit injected at Stratford to reach Huntly. The power flow split for Bunnythorpe is[78]:

---

[78]   Note that this is the same as the split which occurs in the diagram of the unconstrained dispatch.

$$\left(\frac{\text{power flow}_{\text{Bunnythorpe-Stratford-Huntly}}}{\text{power flow}_{\text{Bunnythorpe-Huntly}}} \right) = \left(\frac{\text{resistance}_{\text{Bunnythorpe-Huntly}}}{\text{resistance}_{\text{Bunnythorpe-Stratford-Huntly}}} \right)$$

$$= \left(\frac{2}{1+1} \right) = 1$$

This means that, for every megawatt by which generation is decreased at Bunnythorpe, there is a 0.5MW decrease in flow to Huntly via Stratford and a 0.5MW flow decrease on the more direct, but constrained, link.

To meet an increase of 10MW at Huntly therefore, we must increase generation at Stratford by 20MW, increasing flow on the constrained line by 20×0.25=5MW, while decreasing the injection at Bunnythorpe by 10MW, decreasing flow on the constrained line by 10×0.5=5MW, and cancelling out the extra flow caused by Stratford's increase[79]. The total change in cost, over an hour, of the extra demand at Huntly is:

$$(20MW \times \$20/MW) - (10MW \times \$10/MW) = \$300$$

That is, every actual megawatt-hour of demand north of Huntly has cost the system $30, hence the price at Huntly must be $30/MW[80]. This solution is shown in Figure 5.9.
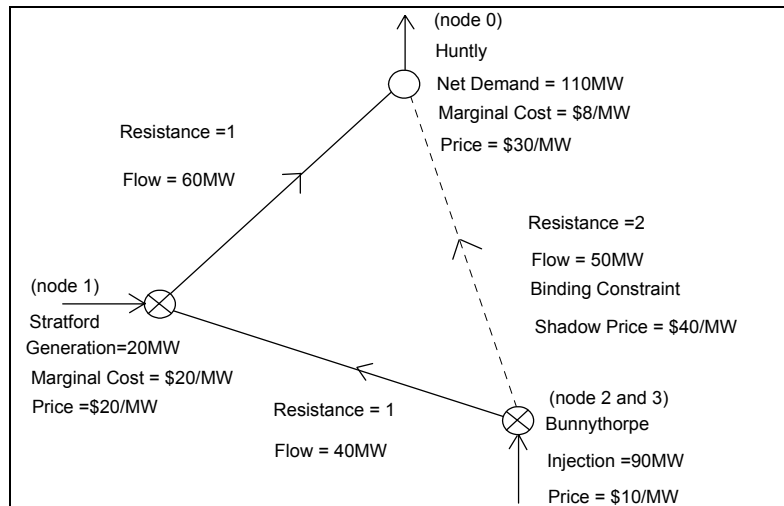
Figure 5.9: A constrained three node network.

---

[79] It may appear that a cheaper approach would be to increase generation at Bunnythorpe by 10MW while decreasing net generation (by increasing demand) at Stratford by 20MW. While this would cause the change in flow on the constrained line to be zero the total change in generation would actually be -10MW, and hence demand would have to be reduced, rather than increased, at Huntly.

[80] Naturally, if there were a generating station north of Huntly with a price greater than $20/MW but less than $30/MW, it would be cheaper to use that to meet the increased demand instead of Stratford at $20/MW.

The value of capacity on the constrained line can be determined by evaluating the change in the cost of supplying power to Huntly if one more unit of capacity were available. This change in cost must reflect the change in the value of the injection at the two marginal nodes. We have already shown that every extra megawatt of power injected at Bunnythorpe increases the flow across the constraint by 0.5MW, while each extra megawatt at Stratford increases this flow by 0.25MW. Thus if an extra unit of capacity were to become available we could change the marginal generation levels according to:

$$0.5\Delta P_{\text{Bunnythorpe}} + 0.25\Delta P_{\text{Stratford}} = 1$$

But we also require that power be conserved. Since we are ignoring losses this power conservation condition is just:

$$\Delta P_{\text{Bunnythorpe}} + \Delta P_{\text{Stratford}} = 0$$

The cheapest solution to these two equations is[81]:

$$\Delta P_{\text{Stratford}} = -4\text{MW}$$
$$\Delta P_{\text{Bunnythorpe}} = +4\text{MW}$$

The resulting change in system cost is:

$$(-4\text{MW}\times\$20/\text{MW}) + (4\text{MW}\times\$10\text{MW}) = -\$40$$

Hence the shadow price on the constraint is \$40/MW[82], reflecting the fact that an extra unit of capacity will reduce the cost of running the system by \$40. Thus we should be prepared to pay \$40 for an extra megawatt of line capacity in the direction of transfer. Note that, as explained above, this shadow price is greater than the price difference between the two ends of the constrained line which, in this case, is only (30-10)=\$20. This reflects the fact that the expansion of the line is not solely for the benefit of those wanting to transfer power directly between these two nodes, but also for those whose transfers indirectly, or partially, utilise that link.

Conversely, this reflects the fact that even a "direct" transfer between these two points will not be solely carried by the constrained line. In fact, in this case, only half is carried by that line, so that the congestion rental for such transfers is only half the \$40/MW shadow price, that is, \$20/MW of additional demand at Huntly.

---

[81]   We could equally reverse the signs, but this would mean increasing generation at the most expensive generating station (Stratford), which is not economically desirable.

[82]   Strictly, given that we are dealing with a cost minimisation problem, this should be -\$40/MW.

5.4.   More Complex Loop Structures

The Spring Washer effect is the simplest case of a more general class of pricing phenomena, involving multiple constrained lines in a single loop, or multiple connected loops with one or more constrained lines. The nature of the loop, and the number and positions of the constrained lines, as well as the direction of the constrained flows, all have implications for the number and position of the marginal nodes. In some cases, many marginal nodes may need to work in unison to get power across a constrained line, while, in other cases, it may be physically impossible to transfer power between two regions, so that completely independent marginal nodes are required in each region.

While the pricing behaviour around such intricate loops may be quite complex, prices will still obey the basic principles underlying the spring washer effect. So, for example, where branches are connected to a single constrained loop in such a way as to form parallel loops, prices will rise around those parallel loops in a similar manner.

5.5.   Conclusions

In this chapter we have extended the general dispatch based pricing model of Chapter 4 so that it includes the impact of transmission constraints, whether they be simple flow constraints or thermal constraints. The form of the dual pricing constraints are similar to those derived by Hogan (1991). We have explained the basic economic impact of these constraints in terms of standard duality theory.

An analysis of prices produced by Trans Power (NZ) Limited's AC pricing model demonstrates the standard effects, as often reported in the literature, of transmission constraints on acyclic transmission lines. However, for constraints in cyclic loops it is shown that a more complicated "spring washer" effect occurs as a consequence of the need to maintain a constant power flow across the constrained line. Two marginal nodes are required to react simultaneously so as to satisfy this constraint. As these nodes will have different prices the prices around the network are forced to rise from the side of the constraint nearest the cheapest marginal node around to side nearest the most expensive marginal node.

We have demonstrated that the spring washer effect is observed in reality, using prices observed in the central North Island of New Zealand, and have presented (illustrative) pricing equations, based on a DC approximation, which are consistent with the behaviour observed in practice. The spring washer effect we describe is fundamentally different from that reported by Wu et al. (1994) as we assume an optimal primal dispatch, with nodal prices and transmission rentals required to be consistent.

Chapter 6

Reserve Capacity Constraints[83]

6.1. Introduction

Electricity generation and transmission equipment are subject to occasional, though generally infrequent, failures. These failures can lead to a disruption of supply in all or part of a power system. These power disruptions place greater loads on the generating units which remain in operation. In New Zealand, this greater load immediately causes the system frequency to drop below its nominal value of 50Hz. Frequencies of AC power which are excessively low or excessively high may damage components of the power system and even the equipment of consumers. Furthermore, New Zealand law defines clear limits as to how far the steady state system frequency can deviate from 50Hz. Consequently, the New Zealand system is operated under the requirement that reserve capacity must be available, both to prevent the system frequency dropping too low, and to quickly restore the frequency to near its normal level.

In a regulated environment, or an environment with centralised decision making, it is relatively easy to manage the provision of reserve. However, in a deregulated or privatised market, reserve provision becomes more difficult to manage. Individual generating companies will tend to have small collections of plant, and will not generally be able to cover their own reserve needs efficiently. Furthermore, no generating company or consumer will want to forego profitable and reliable participation in the energy market to play a reserve role unless adequately compensated. More fundamentally, the dispatch based pricing approach relies on being able to determine a set of prices which is consistent with the observed dispatch. If the dispatch is significantly affected by a binding reserve constraint, it must be accounted for in the pricing system. In countries with small power systems, like New Zealand, the loss of a single generating unit could account for a significant proportion of total generation, making incentives for reserve availability crucial. In fact, for the New Zealand system, the reserve requirement is probably the

---

[83]   This chapter is based on Ring, Read, and Drayton (1993a, 1993b), and Ring and Read (1994c).

single most important requirement, apart from the fundamental energy balance equation, in determining spot prices.

The literature on reserve pricing and markets has been relatively limited. Stadlin (1971) used a Lagrangien multiplier to represent the marginal impact of a generating unit's regulating margin, that is the amount a generating unit can ramp up by in a defined time interval, on system operating costs. This relationship was developed for scheduling reserve rather than for operating a reserve market. A similar idea is adopted in model of Caramanis et al. (1987). This model, while providing a good economic framework, is rather general and does not address many of the complications encountered in reality. Berger and Schweppe (1989) propose the use of MIT style price signals to guide a system through the critical phase immediately following an outage[84]. Hogan (1992) briefly describes simple ways of modifying *ex post* prices to account for transmission contingencies and power outages but does not explicitly model the reserve generation problem. The potential for futures markets in reserve capacity is described by Ruff (1992).

In this chapter we develop a form of reserve pricing which differs from earlier works in a number of respects. We base our approach upon the current contingency planning rules employed in the New Zealand system, and use realistic representations of the behaviour of the reserve sources in that system. This model is deterministic, as the New Zealand system is dispatched so as to be able to cover the single largest contingency, without accounting for the probability of failures occurring[85]. The possibility that reserve may actually be called upon is ignored, on the assumption that this will only happen very infrequently[86]. While not necessarily consistent with the idea of minimising the expected cost of outages, this approach reflects the importance of reserve capacity in the New Zealand system. A discussion of the stochastic aspects of this problem is presented in Chapter 9.

---

[84]   The models of Caramanis et al. (1987) and Berger and Schweppe (1989) are stochastic models and are discussed further in Chapter 9.

[85]   This is principally due to the requirement to remain above a minimum frequency at all times, rather than maintaining a "loss of load probability" requirement as is more usual in larger power systems.

[86]   This approach can be viewed as a regime under which market participants see an approximation to an expected reserve capacity price, and receive no additional compensation (apart from payments for the extra energy, at the energy price) when reserve is called upon. In general, though, the marginal cost of production during contingency events is actually less than before the contingency, as cheaper power sources, which were held back for reserve purposes, are brought into play, and these will make a profit at the prevailing energy price.

A detailed description of the reserve problem as it relates to the New Zealand power system is provided in the following section. In Section 6.3 we explore and analyse a dispatch based pricing model for the pricing of reserve provided by generating units.

6.2. The Reserve Problem in New Zealand

The range of contingencies which can occur in a power system is wide, and many less frequent contingencies cannot be covered without reducing load. In New Zealand, operational guidelines define what sort of contingencies the system should be able to handle without interrupting service. These are essentially rule-of-thumb guidelines, and are not necessarily economically optimal in a global sense. The power system is expected to be able to cover the single worst possible contingency. That is, there must be enough reserve generating capacity in the system, or the systems of each of New Zealand's two main islands, to cover the loss of the largest single source of power. This is known as the "N-1" contingency requirement, where there are N generating units. Similarly, the transmission system is expected, where physically possible, to be able to cope with the loss of a major transmission link.

Until the recent expansion of the high voltage direct current (HVDC) inter-island link, the maximum nominal power level to be covered by spinning reserve in the North Island was usually the loss of one of the four Huntly machines (a maximum of 250MW). With South Island generation usually cheaper, on the margin, than that in the North Island, the HVDC transfer will often be maximised by the dispatcher. Thus the loss of one pole (or line) of the HVDC link (a maximum of 550MW), is now generally the contingency that needs to be covered by North Island spinning reserve.

The largest risk covered in the South Island is the loss of a Clyde machine (a maximum of 108MW). South transfer by the HVDC link is rarely large enough to cause the South Island to have to cover the risk of its loss[87].

6.2.1. The Sources of Reserve

The main form of reserve capacity is that carried by generating units which are producing active power below their full capacity, but which can ramp up their output

---

[87]   One exception being during the 1992 power shortage, when the levels of the South Island hydro lakes were unusually low. The HVDC link was, at times, running with a high transfer south, so requiring that the South Island hydro system carry reserve in excess of the power being generated there.

rapidly if required to. We refer to this reserve as *spinning reserve*. This form of reserve is relatively cheap and readily available, though the provision of spinning reserve generally means that the cost of the dispatch rises due to there being more machines running.

Another form of generating unit reserve is provided by stand-by generating units. These units are not operating at all when an emergency occurs, but can be brought on-line rapidly afterwards. The role of stand-by units is to reduce the stress on the system once the system has recovered from the initial stages of an emergency.

An alternative to increasing generation is to decrease load. Load curtailment involves the voluntary shedding of all, or part, of a consumers load. The amount which can be dropped, the warning times to be given, and the compensation required can be defined by a contract. A potential difficulty in pricing curtailable load is that a customer's demand for power may rise in discrete steps. For example, a factory may have a number of production lines and, if any load is to be dropped, an integer number of production lines must be shut down. This integer aspect could preclude the usage of pure optimal marginal cost pricing techniques. Oren and Smith (1992) describe the load curtailment problem, as well as a simple approach to scheduling and pricing curtailable load given limits on the number of allowable interruptions per year.

The reserve source of final resort is blackout. Blackout is similar to load curtailment but consumers of power have no choice as to whether to accept loss of service. By definition, the marginal cost of blackout equals the local shortage cost.

Transmission capacity can also be viewed as a form of reserve, both because it allows generating units to provide reserve to distant locations, and because spare transmission capacity may be required to cover transmission line failures. Scheduling of reserve transmission capacity should be easier than generation capacity as we assume centralised operational control of the transmission network.

6.2.2. The Reserve Constraints

The reserve constraints, as applicable to New Zealand, are described by Miller and Turner (1991). The approach of Miller and Turner assumes an "N-1" contingency scenario which can be covered through the use of spinning reserve and a limited amount of curtailable load.

When a disruption occurs the system frequency will generally fall below its nominal level of 50Hz. Figure 6.1 shows the typical behaviour of the frequency following a disruption. The system frequency will generally reach a minimum within five to seven

seconds. It is important, however, that this frequency not fall below 48Hz[88]. This constraint is called the *Minimum Frequency Constraint*. The operating generating units must increase their outputs under governor control within the five to seven second interval to ensure that this minimum frequency is maintained.
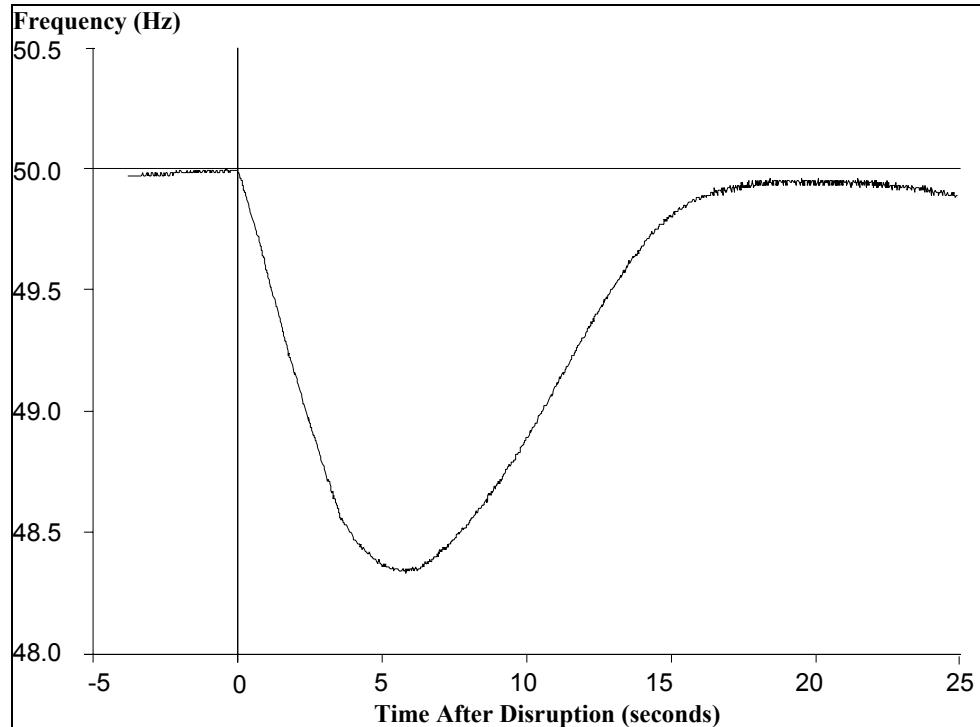
Figure 6.1: North Island frequency response to the failure of one unit at Huntly.

In addition to the frequency response requirement, there must be sufficient surplus reserve power available to permit the system to generate power close to 50Hz once the system returns to an equilibrium state. This is the *Surplus Reserve Constraint*. This is defined in New Zealand by requiring that there be some non-negative amount of reserve generation unused once the system has returned to 49.5Hz. This surplus reserve is measured once the frequency has returned to 49.5Hz, approximately 20 to 30 seconds after the disruption. The longer term issue of returning the frequency to 50Hz is ignored here, but is briefly discussed in Chapter 9.

There is a strong correlation between the surplus reserve and the minimum frequency constraints. This arises because these functions are both measures of the system's ability to contribute "instantaneous reserve". However, while the surplus reserve function is measured in terms of active power, the minimum frequency function is measured in terms

---

[88]   The value of 48Hz is a limit set by Trans Power (NZ) Limited and corresponds to a 0.50Hz safety margin above the frequency at which progressive load shedding (blackout) occurs.

of frequency. The functions defining the surplus reserve and the minimum frequency constraints are evaluated by numerical, rather than analytical, techniques using the model of Miller and Turner (1991). In practice, the minimum frequency constraint is often the only binding constraint.

The speed with which an in-service generating unit can increase its output (utilise its spinning reserve) varies between units. This means that the choice of dispatched generating units and their level of generation, depends not only on the availability and cost of power, but also on the marginal impact each unit can make to the system frequency and total generation during the critical period after a disruption has occurred.

6.2.3. The Reserve Contributions of Generating Units

If we are to price accurately for reserve capacity we must have an understanding of the relationship between a generating unit's reserve contribution and its operating level. The ability of generating units to provide spinning reserve vary significantly with machine loading and machine type. The discussion in this section is illustrated by the general reserve response curves shown in Figure 6.2[89]. The vertical axis of these graphs describe the contribution of a generating unit towards satisfying either the minimum frequency or surplus reserve constraint.

---

[89] These curves, while based on results from the work of Miller and Turner, are purely illustrative examples and do not necessarily correspond to any actual individual generating unit.
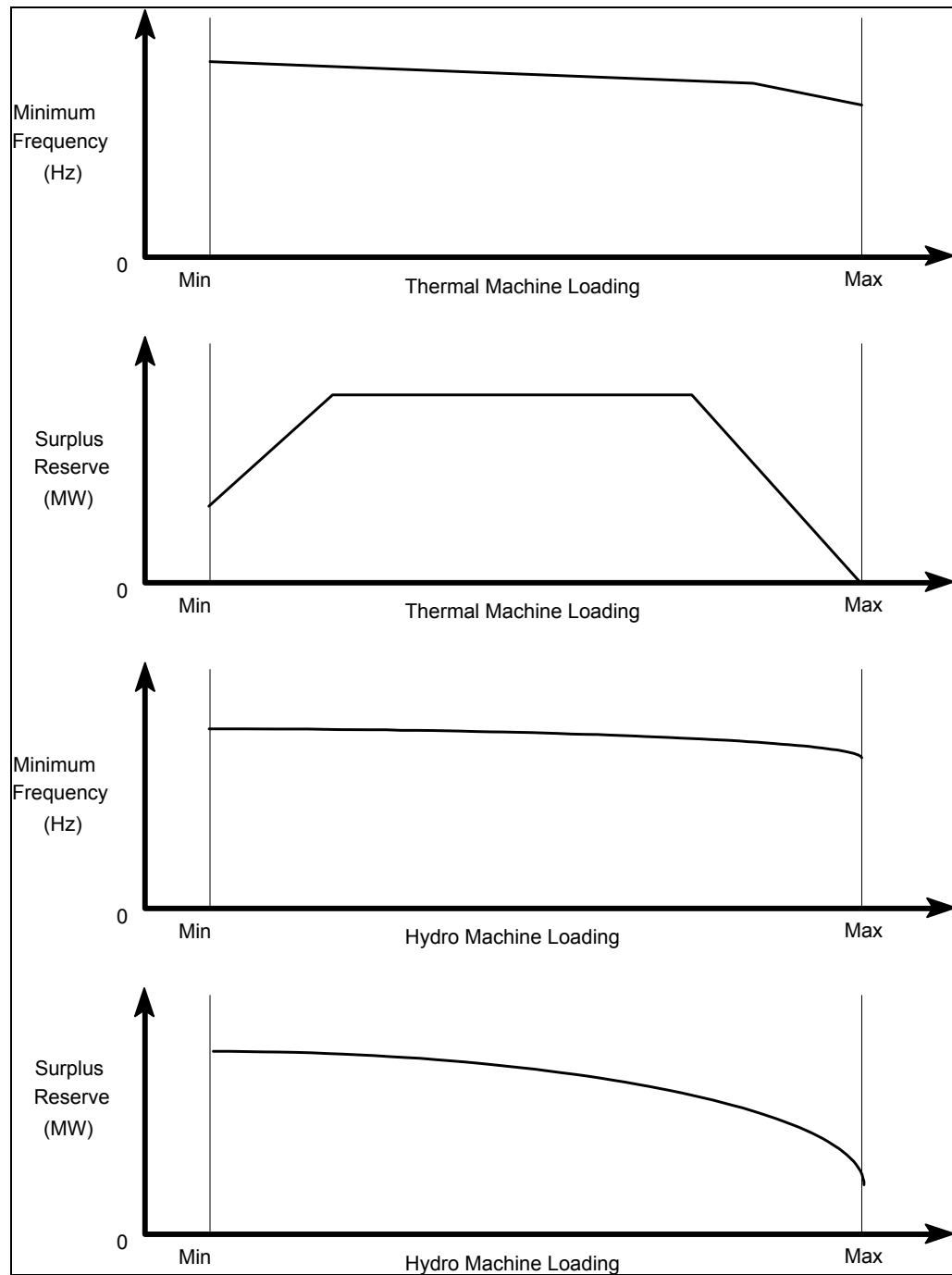
Figure 6.2: Typical reserve contribution functions of thermal and hydro machines.

The form of a thermal generating unit's contribution to the minimum frequency and surplus reserve requirements differ greatly. A thermal generating unit's contribution to the minimum frequency requirement declines with increasing generation over the entire generating range. In contrast, the relationship between generation level and contribution to the surplus reserve requirements can be divided into three ranges. When a thermal generating unit is operating at low output the dynamic response of its boiler means it can

respond only minimally to a sudden increase in load. For this reason the surplus reserve capability for low generation levels is low but increasing with generation (In practice, the generating unit may not provide any reserve until it is at some minimum output level.)[90]. The maximum reserve is available over an intermediate range. In this range it is possible to increase generation by the maximum amount within the time required to meet an increase in net load. At high generation levels the ability of the unit to increase output becomes restricted by the maximum loading limit. Reserve decreases above this level, and none remains when the unit reaches its full capacity. This latter situation is assumed in many simplistic analyses of spinning reserve in which each extra unit generated is assumed to reduce the reserve by approximately one unit.

The forms of the thermal generating unit reserve curves we have described are not comprehensive. In particular, a thermal generating unit is capable of building up a head of steam, that is, operating in "fixed pressure" mode instead of the more economical "sliding pressure" mode. This reduces the efficiency of normal operations, but increases that unit's ability to supply reserve. Consequently, the unit's reserve contribution function becomes dependent on two variables rather than one. As this is not a major issue we ignore it here.

Hydro generating units behave differently from thermal generating units. When a partially loaded hydro unit is subjected to additional load, output drops momentarily with the inertia of increased water flow, before picking up to provide extra power. Partially loaded hydro machines provide additional power at a slower rate than thermal generating units, but, once a hydro generating unit's output is increased, it can maintain the higher level more easily than thermal units. Hydro operating loading levels are fairly strictly governed. This is because a typical hydro machine should not, in general, be run at less than 50% loading, both for operational efficiency and to avoid possible cavitation damage to the turbine. Consequently, a hydro units ability to contribute to meeting reserve requirements may be defined over a smaller range than for thermal generating units (generally 50%-90% loading).

Hydro generating units can also be run in *tail-water depressed* mode. In this mode the unit operates as a motor, powered by the system, with the turbine spinning in air. Hydro machines running in tail-water depressed mode respond by starting to generate when the system frequency falls to a "trip frequency" less than 50Hz, typically 49.0Hz.

---

[90]   This feature of the reserve function may also act as means of modelling integer aspects of the dispatch, at least for dispatch purposes, if not pricing. If a generating unit has a non-zero, positive minimum operating level, and a non-zero reserve contribution at this point, inserting an upward sloping segment, from the origin, to that point, yields a continuous problem, while discouraging generation in the forbidden range.

Upon being tripped, water is allowed to flow past the turbine. The rate at which additional power is supplied to the system is less than that of partially loaded hydro machines due to the time taken for the system frequency to fall to the trip frequency, and the time taken for the stationary water to be accelerated through the turbine.

Figure 6.2 shows the typical form of a hydro generating unit's, or a hydro catchment's, contribution to the two reserve requirements. In both cases the ability to provide reserve falls off with increasing loading.

6.2.4. Reserve and the Transmission System

The effectiveness of reserve capacity provided by generating units depends on the transmission system being able to cope with the altered power flows during a contingency. The transmission system also has to provide reserve capacity to cover line failures. This means that the power flow equations must be feasible for both the pre-contingency and post-contingency states of the system. This approach, which is used by Hogan (1992), is described in more detail in Chapter 9. If insufficient transmission capacity is available, then additional spinning reserve, load curtailment or, in extreme cases, blackouts may be used in the region which loses power.

The task of providing reserve is potentially complicated by losses in the transmission network. If a generating unit provides some level of reserve at one point in the network then it can supply a greater or lesser quantity to other points in the network. In theory, this spatial aspect may significantly complicate the reserve problem. In practice it may be ignored, or greatly simplified, because the New Zealand reserve requirements are seldom calculated to such a fine level of detail that variations due to losses become important.

A link like New Zealand's inter-island HVDC link requires special attention as it can be the largest single source of power in the receiving region. In practice, the HVDC link can be modelled as a generating unit in the receiving island and as a load in the sending island.

6.3. Pricing for Spinning Reserve

6.3.1. Introduction

The observed costs of providing reserve can be used in determining dispatch based prices for generation and reserve. These prices must give generating companies an incentive to provide reserve. In the absence of binding reserve and other operational

constraints, the least cost dispatch to meet the load would always be determined by dispatching stations according to the merit order, with the marginal cost of operating the most expensive, or marginal station, defining the marginal value of supplying power. In theory, it could be possible for the optimal merit order dispatch to provide more than enough reserve to meet any reserve requirements. If so, the reserve constraints are slack and reserve can be provided at zero cost. In general, at least one reserve constraint (minimum frequency or surplus reserve) will be binding, so it will be necessary to alter the dispatch so as to provide this reserve. This requires spreading the load across more generating units, some of which must operate below their full capacity. This out of merit order dispatch will force the cost of supplying power to rise.

In this section we present a mathematical representation of the reserve problem as it applies to the New Zealand system. We will consider only reserve provided by generating units, and assume it to be dependent only on active power generation, as it is the predominant form of reserve. Curtailable load and blackouts can be modelled in an analogous fashion to spinning reserve. We present a dispatch based pricing model for spinning reserve, deriving it from the general model of Chapter 4, and explore the economic interpretation of these prices.

6.3.2. The Primal Reserve Constraints

A set of reserve constraints, *RC*, can be represented as:

$$R_h(\boldsymbol{P_G^{PXS}}) \geq 0 \qquad\qquad \forall h \in RC \quad (6.1)$$

Here $\boldsymbol{P_G^{PXS}}$ is the standard vector of active power generation levels. Experience in New Zealand demonstrates that the contribution of a generating unit to a given reserve constraint is described by a concave function. For simplicity we assume a general form of this function as shown in Figure 6.3. The precise form and number of sections making up this function for a given generating unit depend upon the characteristics of that unit.
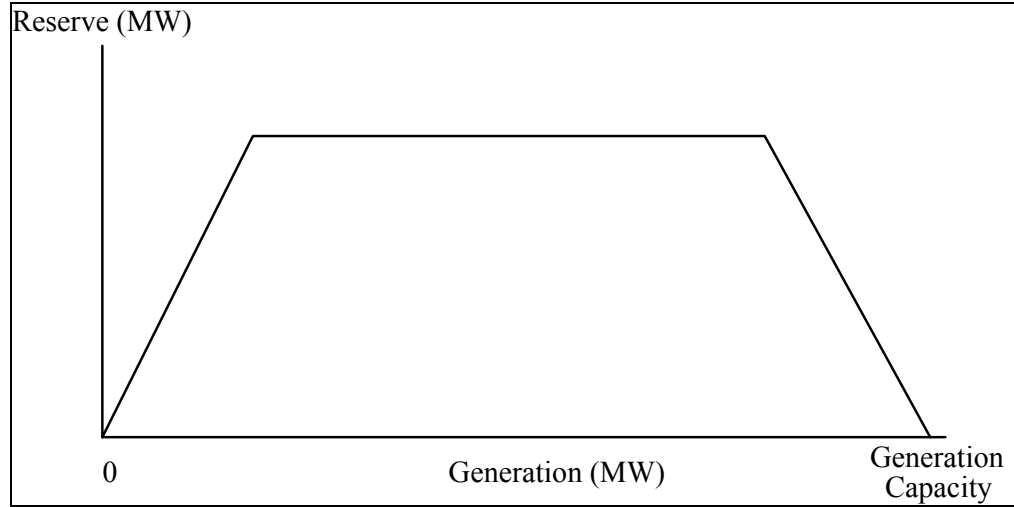
Figure 6.3: A general representative generating unit reserve availability curve.

Given the output levels of all the generating units, we can estimate the marginal contribution of any unit to a reserve requirement[91]. In doing this we are not estimating the entire reserve function, rather, we are estimating the derivatives of the reserve function with respect to the generating unit's output level. We can define ρ_{hi}^- to be the derivative of reserve constraint h with respect to the last unit of active power generated at node i[92], and ρ_{hi}^+ to be the derivative of reserve constraint h with respect to the next unit of power that could be generated at node i[93]. These distinctions are required to account for the discontinuity of the reserve constraint derivatives at the corner points in Figure6.3. Note that convexity of the reserve function implies that $\rho_{hi}^- \geq \rho_{hi}^+$ for all i and h. If P_{Gi}^* is the observed level of generation at node i then we can define the linearised reserve contributions of generating unit i to be:

$$R_{hi}(P_{Gi}) = R_{hi}(P_{Gi}^*) - \rho_{hi}^- P_{Gi}^- + \rho_{hi}^+ P_{Gi}^+ \qquad \forall h \in RC \quad (6.2)$$

---

[91] Tail-water depressed plant is a potential exception. When called upon in an emergency, these generating units switch from consuming power to producing power, the power consumed invariably bearing little relation to the power that can be produced. As this is an integer relationship it is not strictly possible to define their marginal contribution to the system reserve requirements. In practice, however, a satisfactory approximation can be achieved by ensuring that their (approximated) marginal contributions are consistent with the relative values of their generation and reserve contributions.

[92] For ease of discussion we assume that there is only one generating unit per node. Hence generating unit i is the unit at node i. In general, though, we would need to model the contribution of either each individual generating unit or the combined contribution of all generating units attached at a bus-bar.

[93] For a surplus reserve constraint these derivatives have units of MW/MW (i.e., it is unitless), while for a minimum frequency constraint the units are Hz/MW.

Here $R_{hi}\left(P_{Gi}^{*}\right)$ is the reserve contribution at the observed generating level, P_{Gi}^{*}, of the generating unit at node i, while P_{Gi}^{-} and P_{Gi}^{+} represent, respectively, an incremental decrease and increase in generation relative to P_{Gi}^{*}. For later reference, we observe that $R_{hi}\left(P_{Gi}^{*}\right)$ can also be represented as:

$$R_{hi}\left(P_{Gi}^{*}\right) = a_{hi} + \rho_{hi}^{-} P_{Gi}^{*} \qquad\qquad \forall h \in RC \quad (6.3)$$

Here a_{hi} is a linearisation constant.

The combined first-order reserve contributions to constraint h by all the generating units will not necessarily match the total reserve requirement. This difference reflects the fact that the reserve constraints may not be additively separable. Consequently a linearised version of (6.1) might be represented as:

$$R_{h}\left(\boldsymbol{P_G}\right) = R_{h0} + \sum_{i \in PXS} R_{hi}\left(P_{Gi}\right) \geq 0 \qquad\qquad \forall h \in RC \quad (6.4)$$

Here R_{h0}, which may be positive or negative, describes that fraction of the reserve provided which cannot be attributed to any particular generating unit. The physical units of R_{h0} would be megawatts (MW), in the case of the surplus reserve constraint, and hertz (Hz), for the minimum frequency constraint. Substituting (6.2) into (6.4), and moving the constants to the right-hand side, gives:

$$\sum_{i \in PXS}\left(-\rho_{hi}^{-} P_{Gi}^{-} + \rho_{hi}^{+} P_{Gi}^{+}\right) \geq -R_{h0} - \sum_{i \in PXS} R_{hi}\left(P_{Gi}^{*}\right) \qquad \forall h \in RC \quad (6.5)$$

A Fortran program called SPIN, developed by Drayton et al. (1992), calculates the reserve derivatives for generating units and the HVDC link. SPIN uses the frequency routines used by Trans Power (NZ) Limited, which are based on the discussion in Miller and Turner (1991). These derivatives are calculated by determining the marginal impact on each reserve constraint of a unit increase, or decrease, in generation at each node. This analysis includes the *risk bus*, the term used to describe the single largest source of power, the failure of which is the contingency being covered. The derivative of the surplus reserve constraint with respect to the generation at the risk bus is typically close to -1, that is, one more unit of generation from the risk bus increases the required surplus reserve by one unit. In general, the derivative of the surplus reserve constraint with respect to the generation of a generating unit supplying reserve ranges from -1 to +1. For the minimum frequency constraints the derivatives are typically of the order of a fraction of one hertz per megawatt.

In addition to these derivatives, the program estimates the actual megawatt contribution made by each reserve source to the surplus reserve constraint and the contribution, in hertz, by each reserve source to the minimum frequency constraint. This is

done by removing each reserve source in turn from the dispatch and recalculating the surplus reserve and minimum frequency constraints. The resulting megawatt change in surplus reserve is called the *surplus reserve contribution* of a source, while the change in minimum frequency is the *minimum frequency contribution*[94]. For the risk bus these contributions are negative, that is, removing the risk from the problem will improve the relevant reserve constraint.

6.3.3. The Dual Reserve Pricing Relationships

Arranging (6.5) in the form of the general primal constraint (4.6) gives:

$$H_h^A\left(\boldsymbol{P_G^{PXS}}\right) = \sum_{i \in PXS}\left(-\rho_{hi}^- P_{Gi}^- + \rho_{hi}^+ P_{Gi}^+\right) + R_{h0} + \sum_{i \in PXS} R_{hi}\left(P_{Gi}^*\right) - \text{slack} = 0 \qquad \forall h \in RC \quad (6.6)$$

Defining the reserve multiplier such that $v_h \equiv \gamma_h$, and substituting the derivatives of (6.6) into (4.15) to (4.23) (constraint (4.24) is redundant) produces the reserve pricing model shown in (6.7) to (6.15). The units of γ_h are \$/MW for a surplus reserve constraint and \$/Hz for a minimum frequency constraint[95].

---

[94]   This calculation is somewhat approximate, but the details do not affect our analysis here.

[95]   The product of γ_h and ρ_{hi}^-, or ρ_{hi}^+, is always in units of \$/MW.

$$\underset{\substack{\boldsymbol{\gamma}^{RC},\boldsymbol{v}_P^{+PXS},\boldsymbol{v}_Q^{+PXS},\boldsymbol{v}_V^{+PXS}\geq\mathbf{0} \\ \boldsymbol{v}_P^{-PXS},\boldsymbol{v}_Q^{-PXS},\boldsymbol{v}_V^{-PXS}\geq\mathbf{0} \\ \lambda_P,\lambda_Q,\boldsymbol{\mu}_Q^{PV},\boldsymbol{\mu}_V^{PQ},\boldsymbol{\beta}_P^{PXS},\boldsymbol{\beta}_Q^{PXS}}}{\textbf{Minimise}} \qquad \text{Arbitrary Objective Function} \tag{6.7}$$

subject to :

$$\beta_{Pi} = \ \lambda_P\left(1 - \frac{\partial L_P^A}{\partial^A P_i}\right) - \lambda_Q\frac{\partial L_Q^A}{\partial^A P_i} - \sum_{n\in PV}\mu_{Qn}\frac{\partial Q_n^A}{\partial^A P_i} - \sum_{n\in PQ}\mu_{Vn}\frac{\partial V_n^A}{\partial^A P_i} \qquad \forall i\in PXS \quad (6.8)$$

$$\beta_{Qi} = \ -\lambda_P\frac{\partial L_P^A}{\partial^A Q_i} + \lambda_Q\left(1 - \frac{\partial L_Q^A}{\partial^A Q_i}\right) - \sum_{n\in PV}\mu_{Qn}\frac{\partial Q_n^A}{\partial^A Q_i} - \sum_{n\in PQ}\mu_{Vn}\frac{\partial V_n^A}{\partial^A Q_i} \qquad \forall i\in PQS \quad (6.9)$$

$$\beta_{Vi} = -\lambda_P\frac{\partial L_P^A}{\partial^A V_i} - \lambda_Q\frac{\partial L_Q^A}{\partial^A V_i} - \sum_{n\in PV}\mu_{Qn}\frac{\partial Q_n^A}{\partial^A V_i} - \sum_{n\in PQ}\mu_{Vn}\frac{\partial V_n^A}{\partial^A V_i} \qquad \forall i\in PVS \quad (6.10)$$

$$\mu_{Qi} = \ \beta_{Qi} - \lambda_Q \qquad \forall i\in PV \quad (6.11)$$

$$\mu_{Vi} = \ \beta_{Vi} \qquad \forall i\in PQ \quad (6.12)$$

$$c_{Pi}^- - \left\langle\upsilon_{Pi}^-\right\rangle - \sum_{h\in RC}\left\langle\gamma_h\right\rangle\left\langle\rho_{hi}^-\right\rangle \leq \ \beta_{Pi} \leq c_{Pi}^+ + \left\langle\upsilon_{Pi}^+\right\rangle - \sum_{h\in RC}\left\langle\gamma_h\right\rangle\left\langle\rho_{hi}^+\right\rangle \qquad \forall i\in PXS \quad (6.13)$$

$$c_{Qi}^- - \left\langle\upsilon_{Qi}^-\right\rangle \leq \ \beta_{Qi} \leq c_{Qi}^+ + \left\langle\upsilon_{Qi}^+\right\rangle \qquad \forall i\in PXS \quad (6.14)$$

$$\beta_{Vi} = \left\langle\upsilon_{Vi}^+\right\rangle - \left\langle\upsilon_{Vi}^-\right\rangle \qquad \forall i\in PXS \quad (6.15)$$

The key impact of reserve requirements on the pricing problem is due to the reserve terms in (6.13). If we assume that only one reserve constraint is binding (hence making the index h redundant), no generating unit capacity constraints are binding, and $c_{Pi}^- = c_{Pi}^+ = c_{Pi}$, then (6.13) reduces to:

$$c_{Pi} - \gamma\rho_i^- \leq \beta_{Pi} \leq c_{Pi} - \gamma\rho_i^+ \qquad\qquad \forall i \in PXS \quad (6.16)$$

This equation is a more general form of that proposed by Stadlin (1971). If an observed dispatch does not satisfy this equation for every generating unit then either the system has not been dispatched optimally or not all constraints have been accounted for.

The form of (6.16) arises from the fact that, in the energy market, the owners of generating unit i will earn a profit of $\beta_{Pi} - c_{Pi}$ on the last unit of active power generated. If this generating unit is to reduce its generation so as to provide reserve then its (profit maximising) owners will want to receive at least the same profit on the production forgone. That one unit reduction in active power generation changes the reserve contribution by ρ_i^-, producing a return of γ[96] on each unit of reserve. Hence, for there to be adequate incentive to provide this reserve, we require that $\gamma\rho_i^- \geq -\left(\beta_{Pi} - c_{Pi}\right)$. A similar argument explains the right side of (6.16). The optimal point for a generating unit to operate at will be a point where the total combined profit from generation and reserve can be maximised.

Note that, while there are no direct bounds imposed on the reserve prices in (6.13), these bounds are effectively imposed by the difference between the active power price bounds of the marginal generating units. If any reserve constraint is binding then there will necessarily be more than one marginal station. This reflects the fact that an out-of-merit-order dispatch is required. If one of these marginal stations increases its output to meet an increase in demand, other marginal stations may have to alter their output to ensure that the binding reserve constraints remain satisfied. Suppose that we have observed that there are two "truly"[97] marginal generating units (1 and 2), with different marginal costs, in an optimal dispatch. The reason they have been dispatched this way must be that these two generating units differ in their marginal ability to contribute to reserve requirements. Specifically the cheaper marginal generating unit must be partially loaded because it is a more efficient source of reserve. Otherwise we could find a cheaper

---

[96]   Remembering that we assume that reserve is never used, so no fuel costs need be subtracted from this.

[97]   That is, they are marginal sources of both generation and reserve.

dispatch by transferring generation from the more expensive marginal generating unit to the cheaper one. For generating unit 1 we have[98]:

$$\beta_{P1} = c_{P1} - \gamma p_1 \tag{6.17}$$

Similarly, for generating unit 2, we must have:

$$\beta_{P2} = c_{P2} - \gamma p_2 \tag{6.18}$$

Solving for γ gives:

$$\begin{aligned}\gamma &= \frac{(c_{P1} - c_{P2})}{(\rho_1 - \rho_2)} - \frac{(\beta_{P1} - \beta_{P2})}{(\rho_1 - \rho_2)} \\ &= \frac{(c_{P1} - c_{P2})}{(\rho_1 - \rho_2)} - \frac{TP_{12}}{(\rho_1 - \rho_2)}\end{aligned} \tag{6.19}$$

Here TP_{12} represents the cost of transmitting power between marginal nodes 1 and 2. If TP_{12} were to be much smaller than the difference in marginal costs of the marginal generating units, then (6.19) can be approximated as:

$$\gamma = \frac{(c_{P1} - c_{P2})}{(\rho_1 - \rho_2)} \tag{6.20}$$

In this form, the price of spinning reserve, which must explain the discrepancy between the observed marginal costs, is proportional to that cost difference and inversely proportional to the difference in the marginal ability of the two generating units to contribute to reserve. As the denominator becomes smaller, it becomes more difficult to explain why they are both marginal, and very high spinning reserve prices are then required. In the limit, if they have identical marginal reserve contributions, it becomes impossible to explain this discrepancy and the reserve price becomes infinite. In practice, of course, the impact of the transmission price moderates this effect.

Equations (6.17) and (6.18) also suggest that, unless the reserve derivative is zero for one of the generating units, the energy price will not correspond to the marginal cost of any particular generating unit, and that the energy price can only be determined jointly with the reserve prices.

---

[98] In (6.17) and (6.18) we assume that $\rho_1 = \rho_1^- = \rho_1^+$ and $\rho_2 = \rho_2^- = \rho_2^+$.

6.3.4. Profitability of Reserve Provision

The discussion thus far has been based on the marginal value of reserve and generation, but now we look at the total profits implied by reserve pricing. We define the total profit as:

$$\Pi_i = (\beta_{Pi} - c_{Pi})P_{Gi} + \gamma R_i(P_{Gi}) \qquad (6.21)$$

Here $R_i(P_{Gi})$ is the reserve contribution of generating unit i given only one reserve constraint. Substituting for $R_i(P_{Gi})$ from (6.2), and then for $R_i(P_{Gi}^*)$ from (6.3), while setting the changes in the observed generation to zero, gives a profit of:

$$\begin{aligned}
\Pi_i &= (\beta_{Pi} - c_{Pi})P_{Gi}^* + \gamma(a_i + \rho_i^- P_{Gi}^*) \\
&= \gamma a_i + (\beta_{Pi} - c_{Pi} + \rho_i^- \gamma)P_{Gi}^*
\end{aligned} \qquad (6.22)$$

The bracketed term on the second line of (6.22) is the marginal profit earned by generating unit i. If this generating unit is truly marginal then it makes no marginal profit, and the net profit is γa_i (rather than zero as it would be if spinning reserve were not priced). Net profits will exceed this for non-marginal generating units. If generating unit i is a net provider of reserve, then the fact that $R_i(P_{Gi})$ is positive and concave implies that a_i is non-negative, and hence that the owners of generating unit i collect an economic rent on non-marginal units of reserve.

Naturally, a generating unit which makes a negative net contribution to spinning reserve (because a_i is negative or ρ_i^- is negative over too large a range) will have to pay for spinning reserve, and thus receive less under this regime than if spinning reserve were not priced. In the limit, its profits may all be absorbed by reserve payments and it will become marginal, but it will not be forced to operate at a loss. This may often be the case for unit deemed to created the largest contingency[99].

Finally, we must ask whether the system must make a net contribution to, or profit from, spinning reserve prices, or whether it merely effects a transfer between owners of generating units. This depends on whether the reserve constraint is additively separable. If a reserve constraint is additively separable then the sum of the individual generating unit contributions should equal the observed net reserve, which should equal zero, if the

---

[99]   Note that, while the prices which this analysis produces seem valid for generating unit's providing reserve, the attribution of all of these costs to a single unit is not reasonable, in practice, and might create serious anomalies. These anomalies reflect the fact that the standard does not correspond to an economic optimum, and can only be resolved in the context of a more comprehensive stochastic model, as discussed in Chapter 9.

constraint is binding. However, if the reserve constraint is not additively separable, that is, R_{h0} in (6.4) is non-zero, then the system will make a net profit or loss of $\gamma_h R_{h0}$ on reserve constraint h. For example, if the surplus reserve constraint only requires frequency to be returned to 49.5Hz, rather than 50Hz, we should expect to find that the megawatt reserve required to achieve this is consistently less than the megawatt contingency being covered. Thus the system would make a profit out of the reserve market, rewarding the system for the contribution that it makes to meeting the reserve required by being able to recover the frequency reliably after it has reached 49.5Hz.

6.3.5 An Example of Spinning Reserve Pricing

Table 6.1 describes the observed state of three isolated generating units. The reserve constraint has been reduced to the simple requirement that the system can still produce the same output if generating unit 1 were to fail.

i	P_{Gi}^{max} (MW)	P_{Gi}^{*} (MW)	$R_i(P_{Gi}^{*})$ (MW)	ρ_i^{-}	ρ_i^{+}	$c_{Pi}^{-} = c_{Pi}^{+}$ ($/MW)
1	20	20	-20	-1	-1	1
2	25	15	+10	-1	-1	2
3	20	5	+10	$+\dfrac{2}{3}$	$+\dfrac{1}{3}$	3

Table 6.1: Data describing an observed dispatch.

The pricing constraints for the dispatch described by Table 6.1 are:

$$1 - \gamma(-1) \leq \beta_P \leq 1 - \gamma(-1) + \upsilon_{P1}^{+} \tag{6.23}$$

$$2 - \gamma(-1) \leq \beta_P \leq 2 - \gamma(-1) \tag{6.24}$$

$$3 - \gamma\left(+\frac{2}{3}\right) \leq \beta_P \leq 3 - \gamma\left(+\frac{1}{3}\right) \tag{6.25}$$

Because generating unit 1 is fully loaded the term υ_{P1}^{+} in (6.23) can take what ever value is required for pricing consistency. This effectively allows us to ignore the upper price bound in (6.23), so (6.23) simply requires that $\beta_P \geq 1 + \gamma$. Constraint (6.24) requires $\beta_P = 2 + \gamma$ (in units of $/MW) while (6.24) and (6.25) together imply that $0.6 \leq \gamma \leq 0.75$ (in units of $/MW). This result reflects the fact that a unit decrease in demand would be met at least cost by generating unit 2 decreasing generation by 0.4MWh while generating unit 3 decreased generation by 0.6MW. The net change in reserve would be zero and the total decrease in system cost would be $2.6. The combination of prices corresponding to

this is β_P=\$2.6/MW and γ=\$0.6/MW. A unit increase in demand would be met by generating unit 2 increasing output by 0.25MW while generating unit 3 increased output by 0.75MW. Again the net change in reserve would be zero while increasing total system cost by \$2.75. In this case the corresponding prices would be β_P=\$2.75/MW and γ=\$0.75/MW.

We have not specified an objective function for this problem. The objective function which minimises (maximises) the profits of the generating companies amounts to minimising (maximising) β_P. The profits are summarised by Table 6.2.

Maximising Generating Company Profits: β_P=\$2.75/MW, γ=\$0.75/MW.					
i	P_{Gi}^* (MW)	$R_i(P_{Gi}^*)$ (MW)	c_{Pi}^- (\$/MW)	$\left(\beta_P - c_{Pi}^-\right)P_{Gi}^*$ (\$)	$\gamma R_i(P_{Gi}^*)$ (\$)	Π_i (\$)
1	20	-20	1	35	-15	20
2	15	+10	2	11.25	7.5	18.75
3	5	+10	3	-1.25	7.5	6.25
			TOTAL:	45	0:	45
Minimising Generating Company Profits: β_P=\$2.6/MWh, γ=\$0.6/MWh.					
i	P_{Gi}^* (MW)	$R_i(P_{Gi}^*)$ (MW)	c_{Pi}^- (\$/MW)	$\left(\beta_P - c_{Pi}^-\right)P_{Gi}^*$ (\$)	$\gamma R_i(P_{Gi}^*)$ (\$)	Π_i (\$)
1	20	-20	1	32	-12	20
2	15	+10	2	9	6	15
3	5	+10	3	-2	6	4
			TOTAL:	39	0	39

Table 6.2: The prices and profits corresponding to the observed dispatch.

This example demonstrates the importance of pricing for reserve capacity. It is the earnings received from reserve that make it viable for generating unit 3 to operate. With out these earnings either generating unit 3 would not operate, or the price would have to rise to \$3/MW. The former case would lead to a shortage of reserve capacity, the latter would have the consumers subsidising the inadequate reliability of generating unit 1. Charging generating unit 1 for the reserve it requires lowers the costs of power to the consumers.

While, in this example, no generating unit earns less than it would have if no reserve requirements were imposed[100], this is not a general result. For instance, as discussed in Section 6.3.4., the owners of a generating unit may have to use all their profits earned from power generation to cover the cost of the generating unit's reserve requirements.

6.4. Conclusion

In this chapter we have described the nature of the reserve requirements currently employed in the operation of the New Zealand power system, and have presented a dispatch based pricing model corresponding to these requirements. We have demonstrated that the price for reserve capacity availability should fully compensate those generating companies which provide reserve for the opportunity costs incurred on the energy production foregone.

The model discussed in this chapter provide a rather simplistic view of reality, primarily because it ignores the variety of contingencies which may occur, and the way in which reserve may be used to meet them. We have assumed that the costs associated with reserve usage are small due to the low probabilities involved. However, the total expected cost of meeting a contingency may be significant if the cost of actually using reserve is high, as would be the case with blackouts. Furthermore, if a market were to actually operate based on the reserve prices calculated by these methods, with the target reserve required being enough to compensate for the failure of the single largest source of power, then, so as to avoid paying out for reserve, the owner of that source will face an artificial second-order incentive to not be the largest source of power.

Modelling the occurrence of contingencies significantly increases the complexity of the dispatch, and hence the reserve pricing problem. We must model each possible contingency, and the reserve to meet those contingencies including curtailable load and blackout, while simultaneously ensuring that the all the post-contingency power flows are feasible. In Chapter 9 we discuss such a model.

---

[100]   That is, there is no risk of any unit failing, and hence no reserve required. There is, however, more than one sense in which we can define the "absence of reserve requirements". If reserve requirements were ignored by the industry, even with a risk of generating unit failure, then the full cost of these failures would be borne during the contingencies, when the energy price would have to rise dramatically to reflect shortage. In this situation, the only penalty for the owners of a generating unit which fails will be the requirement to buy power to cover their contractual requirements, and all generating units will make positive (marginal) profits in the energy market when they operate, these profits being no less than those for the "no risk" situation.

Chapter 7

A "Best Compromise" Pricing Approach

7.1.   Introduction

The models of Schweppe et al. (1988), Hogan (1991, 1992), and that presented in Chapter 4, determine power prices consistent with an optimal power system dispatch.   In practice however, especially with hindsight, dispatches often appear to be sub-optimal. This can occur because of genuine mis-judgements in the dispatch process, but is more often due to modelling limitations, lack of perfect foresight, and the fact that the system cannot instantly adapt to changing circumstances, or perceptions, as they occur. Mathematically, the fundamental reason for price inconsistencies is that it is impossible to determine a price at node *i* which is consistent with the price at some reference node, taking account of the cost of the marginal losses and congestion charges on binding transmission constraints, while also keeping the price at node *i* between the marginal cost of the last unit of power produced there and that of the next unit of power producible there[101]. The only way the model of Chapter 4 can deal with a sub-optimal primal dispatch variable is to treat the observed value as being constrained, a step which is only practical for commodities not actually traded on the spot market, as the prices associated with these variables may have no relation to the costs.   In general, the model of Chapter 4 cannot determine prices economically consistent with an observed sub-optimal dispatch. Mathematically, this follows from the well-known result that the dual of a sub-optimal primal is not feasible.

Wu et al. (1994) circumvent the problem of primal sub-optimality by abandoning the requirement that inter-nodal price differences be consistent with marginal costs, an approach which allows the cost of any primal sub-optimality to be included in the transmission charges. This approach fails to produce useable marginal cost information about the quality of the dispatch, and is inappropriate for the market structure we consider.

---

[101]   If node *i* is a demand node then the price must lie between the marginal benefit which could be gained from consuming one more unit of power, and the benefit gained from the last unit of power consumed. In practice we can treat consumers as generators of negative power and model them in the same manner.

An alternative approach to the problem is to introduce some slack, which is controlled by the pricing objective function, into the pricing constraints. Such an approach was proposed by Schweppe et al. (1988) to account for a revenue constraint. The revenue constraint was satisfied by minimising the weighted least squares deviations of the prices from the optimal spot price values. Vankatesh et al. (1992) apply a similar approach to determining prices associated with a near optimal OPF solution, minimising the sum of squares of the pricing inconsistencies. The least squares approach provides no economic interpretation for the infeasibility, however, and therefore has the potential to achieve price consistency by forcing parties to operate at a loss.

To achieve price consistency we must allow violations of the price bounds subject to an objective function penalty. If these penalties are arbitrary there will still be no economic interpretation of the dual, but we show in this chapter that it is possible to determine economically consistent values for these penalties. Indeed, these penalties should reflect the economic loss suffered by those whose resources are dispatched sub-optimally relative to the market prices. This idea underlies the goal programming approach which we use to find a "best compromise price vector" which minimises the possibly (notional) side payments required to compensate those whose resources, with hindsight, have been inappropriately scheduled. These compensation payments can be viewed as a penalty on those responsible for the sub-optimal dispatch. Even if such payments are never made they are a legitimate measure of the economic loss suffered by the various parties.

We will show that applying this best compromise pricing approach to a truly linear (or convex piece-wise linear) primal problem actually produces prices consistent with the optimal primal dispatch, even though the observed solution may be sub-optimal. For linear approximations to complex non-linear systems the best compromise pricing approach may not reproduce the optimal prices (if they exist), but will produce prices which satisfy the basic commercial requirements of dispatch based pricing. For instance, for an approximation of a non-linear power system, best compromise pricing can produce prices which minimise the compensation "side payments" required, and ensures that no buyer or seller of power will make a loss on the marginal transactions which occur in the sub-optimal solution. We argue that this method provides a robust and workable approach to resolving the dilemmas and disputes that will inevitably arise in real power systems where substantial sums of money are at stake.

In Section 7.2 we derive a best compromise pricing formulation for a linear primal problem. We examine the behaviour of this formulation, and compare it with standard duality theory and *ex post* pricing. We discuss possible applications of best compromise

pricing to piece-wise linear and more general non-linear problems in Sections 7.4 and 7.5. An illustrative example of best compromise pricing is presented in Section 7.6. A discussion of some of the second order incentives that best compromise pricing might create is presented in Section 7.7.

7.2. Philosophy of Best Compromise Pricing

The underlying philosophical basis of best compromise pricing is that, in a dispatch based pricing regime, the dispatcher must provide some economically consistent explanation of any discrepancy between the actual costs of a dispatch and the perceived "minimum" cost dispatch. Regardless of the precise form of bids and offers of market participants, whether these contain purely marginal cost terms or a combination of commitment and marginal cost terms, there will be a feasible dispatch which has the minimum possible cost associated with it. This minimum cost will not necessarily be the minimum cost associated with a truly competitive market, though, unless the bids and offers represent the true costs faced by the market participants.

The "ideal" pricing regime will be commercially consistent with these bids and offers, given the observed dispatch. In practice, however, the system is often going to be dispatched at a higher cost than this ideal minimum, if simply due to the limits on dispatch technology. The commercial implication of these higher costs is that when the corresponding prices are announced some parties might insist that the dispatch of their assets was not optimal with respect to these prices. Furthermore, consumers are ultimately going to have to pay this cost, and will want it to be minimised.

We propose best compromise pricing as a means for providing an economic explanation of this cost discrepancy which also, potentially, creates an incentive to minimise the discrepancy. In Chapter 3 we argue that the dispatcher should determine prices which provide an economic explanation of an observed dispatch, and which are consistent with the bids and offers of market participants. If it is not possible to achieve this solely with standard market clearing prices then some of these prices must necessarily be inconsistent with the bids and offers. In this event, we can introduce additional price terms, which we refer to as *violation prices*, into the problem. Violation prices are applied, in addition to the *market clearing prices*, to those market transactions which would violate bids or offers when evaluated using the market clearing prices above. The combination of market clearing prices and violation prices can now be consistent with the bids and offers, while the market clearing price itself may not be. In effect, we are still applying the conditions of dispatch based pricing, but are expanding the price set to include prices for

resources used consistently with bids and offers (the "market clearing price") and prices for those resources which are not used consistently (the "violation prices").

The violation prices can be used to define the compensation payments which need to be paid to each market participant so as to produce a total payment consistent with that participant's bid or offer. The revenue required to cover these "compensation payments" should be collected from the market as a whole, as a fee reflecting the cost resulting from a decentralised dispatch, technical limitations, and the cost of uncertainty. These funds might be collected by the dispatcher at the beginning of the year and paid out in the form of compensation payments throughout the year. If the dispatcher were to be allowed to keep any of the fee remaining at the end of the year then this creates an incentive for the dispatcher to minimise the need for compensation payments, either by improving dispatch procedures or by increasing the accuracy of the pricing model. Conversely, if the dispatcher were to be liable to pay any excess if the compensation payments exceed the fees collected at the beginning of the year, the dispatcher will be encouraged to dispatch the system in a manner which keeps the observed dispatch cost as close as possible to the minimum possible cost corresponding to a feasible dispatch[102].

Thus this procedure indicates which aspects of the dispatch/pricing model system are worth improving and provides the dispatcher with an incentive to improve them, while ensuring that the market participants are hedged against the cost of mis-dispatch.

The OPF problem, upon which the spot prices of Chapter 4 have been based, assumes a single period, unchanging dispatch, with all integer terms fixed externally. The important implication of this is that the OPF is only an approximation of the more general dispatch problem, and hence provides only limited information to the pricing problem. One implication of this is that best compromise prices produced by a model based purely on an OPF may make what, in the wider sense, is a truly optimal dispatch appear to be sub-optimal, with compensation being paid. Similarly, best compromise pricing make what is a truly sub-optimal dispatch appear to be optimal, with no compensation being paid. In this chapter we assume that the OPF provides a full description of the dispatch problem, but consider the wider issue of compensation in Chapter 10.

---

[102]   Note that care is required in designing these arrangements so that the dispatcher is rewarded for the benefits delivered from improved dispatch as well as facing the penalties for mis-dispatch. The details of such arrangements are beyond this thesis.

7.3. Best Compromise Pricing for Linear Problems

7.3.1. Derivation of a Best Compromise Pricing Model

To derive best compromise prices we assume a general linear primal of the form shown in (7.1) to (7.4). While actual primal problems may not be linear, we will subsequently show how they might be represented as linear problems.

$$\underset{x}{\textbf{Minimise}} \quad \sum_{\ell \in NDI} c_{x\ell} x_{\ell} \tag{7.1}$$

subject to:

$$\sum_{\ell \in NDI} a_{h\ell} x_{\ell} = b_h \qquad : \varsigma_h \qquad \forall h \in CS \tag{7.2}$$

$$-x_{\ell} \geq -x_{\ell}^{max} \qquad : \upsilon_{x\ell}^{+} \qquad \forall \ell \in NDI \tag{7.3}$$

$$x_{\ell} \geq x_{\ell}^{min} \qquad : \upsilon_{x\ell}^{-} \qquad \forall \ell \in NDI \tag{7.4}$$

The vector x comprises variables from the set *NDI*, comprising all dependent and independent variables. This would include, for example, line flows as well as nodal power injections[103]. All these variables have upper and lower bounds and incur a usage cost of $c_{x\ell}$ (possibly zero). The set *CS* describes the set of all constraints which define the relationships between the independent and dependent variables. Each unit of variable x_{ℓ} consumes $a_{h\ell}$ units of resource h, of which there are b_h units available. The dual of this problem is:

$$\underset{\upsilon_x^-, \upsilon_x^+ \geq 0, \varsigma}{\textbf{Maximise}} \quad \sum_{h \in CS} \varsigma_h b_h + \sum_{\ell \in NDI} \left(\upsilon_{x\ell}^{-} x_{\ell}^{min} - \upsilon_{x\ell}^{+} x_{\ell}^{max} \right) \tag{7.5}$$

subject to:

$$\sum_{h \in CS} a_{h\ell} \varsigma_h - \upsilon_{x\ell}^{+} + \upsilon_{x\ell}^{-} = c_{x\ell} \qquad : x_{\ell} \qquad \forall \ell \in NDI \tag{7.6}$$

We can define an observed solution vector for the primal problem to be x^o. We do not assume that this observed solution is optimal, though it is required to be feasible[104].

---

[103]   The issue of the incentives which arise when the dispatcher is also the grid owner are discussed below.

[104]   There is a real possibility that the primal dispatch problem will appear to be infeasible relative to the limits specified in the model. In practice this situation is resolved by accepting that some constraints will be violated at a cost. As suggested by Stott, Alsac, and Monticelli (1987), we can achieve this by allowing the primal constraints to be violated, but subject to objective function penalties. This effectively removes any feasibility problem. Obviously this modification is not

Given this solution, we want to determine a set of prices which will be "consistent", in some sense, with the marginal costs of the resources employed or available. For a linear primal problem it is not difficult to achieve these goals. We first introduce new variables, which describe potential hypothetical variations in the observed solution, as follows:

$$x_\ell = x_\ell^o + \Delta x_\ell \qquad \qquad \forall \ell \in NDI \quad (7.7)$$

Here Δx_ℓ is unconstrained in sign, and represents the change which could have been applied to x_ℓ^o to reach a given value of x_ℓ. This transformation allows us to express a linear primal problem in a form centred on the observed solution, but allowing for adjustments which could have been required to achieve optimality, for example. This formulation is described by (7.8) to (7.11).

$$\underset{\Delta x}{\textbf{Minimise}} \quad \sum_{\ell \in NDI} c_{x\ell}\left(x_\ell^o + \Delta x_\ell\right) \qquad \qquad (7.8)$$

subject to:

$$\sum_{\ell \in NDI} a_{h\ell}\left(x_\ell^o + \Delta x_\ell\right) = b_h \qquad \qquad \forall h \in CS \quad (7.9)$$

$$-x_\ell^o - \Delta x_\ell \geq -x_\ell^{max} \qquad \qquad \forall \ell \in NDI \quad (7.10)$$

$$x_\ell^o + \Delta x_\ell \geq x_\ell^{min} \qquad \qquad \forall \ell \in NDI \quad (7.11)$$

The requirement of primal feasibility means that:

$$\sum_{\ell \in NDI} a_{h\ell} x_\ell^o = b_h \qquad \qquad \forall \ell \in NDI \quad (7.12)$$

Hence the constant terms in (7.9) cancel. We can also remove the constant terms in (7.8) from the minimisation. The resulting primal problem is described by (7.13) to (7.16), where the shadow prices are noted on the far right.

---

necessary for those constraints which cannot meaningfully be violated, such as the energy conservation constraints.

$$\sum_{\ell \in NDI} c_{x\ell} x_\ell^o + \underset{\Delta x}{\textbf{Minimise}} \quad \sum_{\ell \in NDI} c_{x\ell} \Delta x_\ell \tag{7.13}$$

subject to:

$$\sum_{\ell \in NDI} a_{h\ell} \Delta x_\ell = 0 \qquad\qquad : \varsigma_h \qquad \forall h \in CS \tag{7.14}$$

$$-\Delta x_\ell \geq \left(-x_\ell^{max} + x_\ell^o \right) \qquad : v_{x\ell}^+ \qquad \forall \ell \in NDI \tag{7.15}$$

$$\Delta x_\ell \geq \left(x_\ell^{min} - x_\ell^o \right) \qquad : v_{x\ell}^- \qquad \forall \ell \in NDI \tag{7.16}$$

This new formulation is identical to the original optimisation problem described by (7.1) to (7.4), but with the coordinate system now transformed so that all variables are measured relative to the observed solution. At optimality (7.7) requires:

$$\sum_{\ell \in NDI} c_{x\ell} x_\ell^o + \sum_{\ell \in NDI} c_{x\ell} \Delta x_\ell^* = \sum_{\ell \in NDI} c_{x\ell} x_\ell^* \tag{7.17}$$

The right hand sides of (7.15) and (7.16) describe the capacity of variable x_ℓ which is remaining, or has been used, respectively[105]. For a feasible x_ℓ^o these terms will always be non-negative. The change in resource usage required to accommodate a non-zero Δx_ℓ is described by (7.14).

The dual problem corresponding to either (7.1) to (7.4) or (7.13) to (7.16) is:

$$\sum_{\ell \in NDI} c_{x\ell} x_\ell^o + \underset{v_x^-, v_x^+ \geq 0, \varsigma}{\textbf{Maximise}} \quad \sum_{\ell \in NDI} \left(v_{x\ell}^+ \left(-x_\ell^{max} + x_\ell^o \right) + v_{x\ell}^- \left(x_\ell^{min} - x_\ell^o \right) \right) \tag{7.18}$$

subject to:

$$\sum_{h \in CS} a_{h\ell} \varsigma_h - v_{x\ell}^+ + v_{x\ell}^- = c_{x\ell} \qquad : \Delta x_\ell \qquad \forall \ell \in NDI \tag{7.19}$$

This formulation is identical to (7.5) and (7.6), the dual of the original primal. This can be seen by observing that (7.6) and (7.19) are identical, while (7.18) can be re-arranged as follows:

$$\sum_{\ell \in NDI} c_{x\ell} x_\ell^o + \underset{v_x^-, v_x^+ \geq 0, \varsigma}{\textbf{Maximise}} \quad \sum_{\ell \in NDI} \left(\left(v_{x\ell}^- x_\ell^{min} - v_{x\ell}^+ x_\ell^{max} \right) + x_\ell^o \left(v_{x\ell}^+ - v_{x\ell}^- \right) \right) \tag{7.20}$$

But using (7.19) and (7.12) we get:

---

105   Note that x_ℓ^{min} and x_ℓ^{max} are not necessarily the absolute minimum and maximum value of, say, the output of a given generating unit, but the minimum and maximum levels which can be achieved at a marginal cost of $c_{x\ell}$.

$$\sum_{\ell \in NDI} x_\ell^o \left(v_{x\ell}^+ - v_{x\ell}^- \right) = \sum_{\ell \in NDI} x_\ell^o \left(\sum_{h \in CS} a_{h\ell} \varsigma_h - c_{x\ell} \right)$$

$$= \sum_{\ell \in NDI} b_h \varsigma_h - \sum_{\ell \in NDI} c_{x\ell} x_\ell^o \tag{7.21}$$

Substituting this into (7.20) gives (7.5), the original dual objective function.

Dropping the constants from the objective function in (7.18), changing the sign of this objective function, and defining $\beta_{x\ell} = \sum_{h \in CS} a_{h\ell} \varsigma_h$ we get the formulation shown in (7.22) to (7.24), which we will refer to as the best compromise pricing problem.

$$\underset{v_x^-, v_x^+ \geq 0, \varsigma}{\textbf{Minimise}} \quad \sum_{\ell \in NDI} \left(v_{x\ell}^+ \left(x_\ell^{max} - x_\ell^o \right) + v_{x\ell}^- \left(x_\ell^o - x_\ell^{min} \right) \right) \tag{7.22}$$

subject to:

$$\beta_{x\ell} = \sum_{h \in CS} a_{h\ell} \varsigma_h \qquad\qquad \forall \ell \in NDI \quad (7.23)$$

$$\beta_{x\ell} - v_{x\ell}^+ + v_{x\ell}^- = c_{x\ell} \qquad\qquad \forall \ell \in NDI \quad (7.24)$$

7.3.2. Interpretation of the Best Compromise Pricing Problem

We have shown that the formulation in (7.22) to (7.24) is just the dual of the original, linear primal problem, but described with reference to an arbitrary observed solution, rather than necessarily an optimal solution as in the approach of Hogan et al. (1995). Furthermore, as the coordinate transformation we have applied to the primal is a simple translation, with a change in x_ℓ in the original primal formulation having the same impact on the margin as a change in Δx_ℓ in the translated problem, the marginal resources being priced for are identical in both cases. It follows that the dual of the transformed problem will produce the same shadow prices as the dual of the original primal formulation (or, if the dual of the original primal formulation is degenerate, it will select a solution from the same set of degenerate solutions)[106]. Hence, while appearing to make no reference to the optimal primal solution, the best compromise pricing formulation has all the information it needs to determine the direction in which x^* lies, and can determine shadow prices consistent with the optimal solution of the original primal problem.

---

[106]   If $x_\ell^{min} = x_\ell^o = x_\ell^{max}$ then no term corresponding to x_ℓ appears in (7.22), with (7.24) requires that $\beta_{x\ell}$ be unbounded. This is consistent with the dual of the original primal problem, however, as x_ℓ is fixed in value, and duality theory therefore requires that its shadow price be unbounded.

It is apparent from (7.17) and (7.18), and the requirement that the optimal primal objective function value equals the optimal dual objective function value, that:

$$\sum_{\ell \in NDI}\left(\upsilon_{x\ell}^{+*}\left(x_{\ell}^{max}-x_{\ell}^{o}\right)+\upsilon_{x\ell}^{-*}\left(x_{\ell}^{o}-x_{\ell}^{min}\right)\right)=\sum_{\ell \in NDI}c_{x\ell}x_{\ell}^{o}-\sum_{\ell \in NDI}c_{x\ell}x_{\ell}^{*}, \qquad (7.25)$$

That is, the objective function in (7.22) describes the difference in cost between the observed and optimal solutions. Since the primal problem is a minimisation problem this difference must be non-negative. In particular, the value of (7.22) will be zero if $x^{o}=x^{*}$.

The best compromise pricing constraints are identical to the dual constraints of the original primal problem, though we have expressed (7.23) and (7.24) in a form similar to that of (4.61) to (4.66) in Chapter 4. That is, (7.23) defines the pricing relationships while (7.24) describes the bounds on the prices[107]. An important difference, though, is that the pricing relationships defined by (4.61) to (4.63) clearly distinguish between the prices for primal independent and dependent variables, whereas (7.23) makes no such distinction. Equation (7.24) requires that the multipliers on the primal variable bounds, $\upsilon_{x\ell}^{-}$ and $\upsilon_{x\ell}^{+}$ for all $\ell \in NDI$, must exactly account for the difference between the marginal cost of using a variable and the market price, $\beta_{x\ell}$, associated with that variable. For optimally dispatched variables these terms describe the rents earned on the limited capacity available at that cost, in the same manner as discussed in Chapter 4. For resources which have not been used optimally these rents can be thought of as *violation prices*, which apply to the excess, or shortfall, in resource usage. These rents can take any positive value and hence ensure that a feasible dual solution can be found. Regardless of whether we view $\upsilon_{x\ell}^{-}$ and $\upsilon_{x\ell}^{+}$ as violation prices or capacity rents we still require that, for each ℓ, at most only one of $\upsilon_{x\ell}^{-}$ and $\upsilon_{x\ell}^{+}$ will have a non-zero value, as their can only be value in changing a binding primal capacity constraint corresponding to one of these for any given primal optimal solution.

The best compromise pricing formulation can be thought of as a goal programming problem (Lee, 1972), though with the unusual feature of being applied to a dual programming problem. The goal in this context is to determine prices, $\beta_{x\ell}$, which have the greatest possible degree of consistency with the observed primal dispatch and its costs. Equivalently, we can view the objective as being to minimise the capacity rents foregone due to primal non-optimality. We can illustrate this by using the pricing relationships corresponding to an optimal primal solution, x_{ℓ}^{*}, to show how the best compromise pricing objective function works to minimise the compensation corresponding to an observed solution x_{ℓ}^{o}. Thus:

---

[107] Here (7.24) is an equality, rather than an inequality as in Chapter 4, because we have assumed a single cost applies over the range of each primal variable.

- if $x_\ell^* = x_\ell^{min}$ then $v_{x\ell}^- = c_{x\ell} - \beta_{x\ell} \geq 0$ and $v_{x\ell}^+ = 0$. Thus any $\beta_{x\ell} \leq c_{x\ell}$ is an acceptable price. If, however, $x_\ell^o > x_\ell^{min}$ then the objective function acts to minimise $v_{x\ell}^-\left(x_\ell^o - x_\ell^{min}\right) = \left(c_{x\ell} - \beta_{x\ell}\right)\left(x_\ell^o - x_\ell^{min}\right)$. Thus the objective function works to set $\beta_{x\ell} = c_{x\ell}$ or otherwise will work to determine a value of $\beta_{x\ell} < c_{x\ell}$ which minimises the loss due to over-using variable x_ℓ in a situation when its use is not profitable (subject to trade-offs with other compensation terms).

- Similarly, if $x_\ell^* = x_\ell^{max}$ then $v_{x\ell}^+ = \beta_{x\ell} - c_{x\ell} \geq 0$ and $v_{x\ell}^- = 0$. In this situation any $\beta_{x\ell} \geq c_{x\ell}$ is an acceptable price. If, however, $x_\ell^o < x_\ell^{max}$ then the objective function acts to minimise $v_{x\ell}^+\left(x_\ell^{max} - x_\ell^o\right) = \left(\beta_{x\ell} - c_{x\ell}\right)\left(x_\ell^{max} - x_\ell^o\right)$. Thus the objective function works to set $\beta_{x\ell} = c_{x\ell}$ or otherwise will work to determine a value of $\beta_{x\ell} > c_{x\ell}$ which minimises the loss due to under-using variable x_ℓ in a situation when its use is profitable (subject to trade-offs with other compensation terms).

- if $x_\ell^{min} < x_\ell^* < x_\ell^{max}$ then $v_{x\ell}^+ = 0$ and $v_{x\ell}^- = 0$ so no terms relating to x_ℓ have any impact on the best compromise pricing objective function (at its optimal solution). That is, variable x_ℓ is marginal, with $\beta_{x\ell} = c_{x\ell}$, and no profit is made on this variable, regardless of whether it is optimally dispatched or not.

Thus our model generalises the original dual formulation, in which constraints were priced if, and only if, they were binding, by assigning shadow prices to other constraints where necessary. The best compromise pricing objective function tends to assign prices to those constraints which are "most nearly binding", in that x_ℓ^o is close to either x_ℓ^{min} or x_ℓ^{max}, with the trade-off being determined via (7.23). The net effect of this trade-off is to produce shadow prices consistent with an optimal primal solution.

In a market situation, if <u>all</u>[108] suppliers of services are to be paid the optimal prices, then a sum of money, equalling the optimal value of the best compromise pricing objective function, must be injected into the market. This injection can be interpreted as a compensation payment.

---

[108] We discuss below, for the situation where the dispatcher is also the grid owner, a "notional unconstrained network" where capacity provided by the transmission grid owner are excluded from the best compromise pricing formulation.
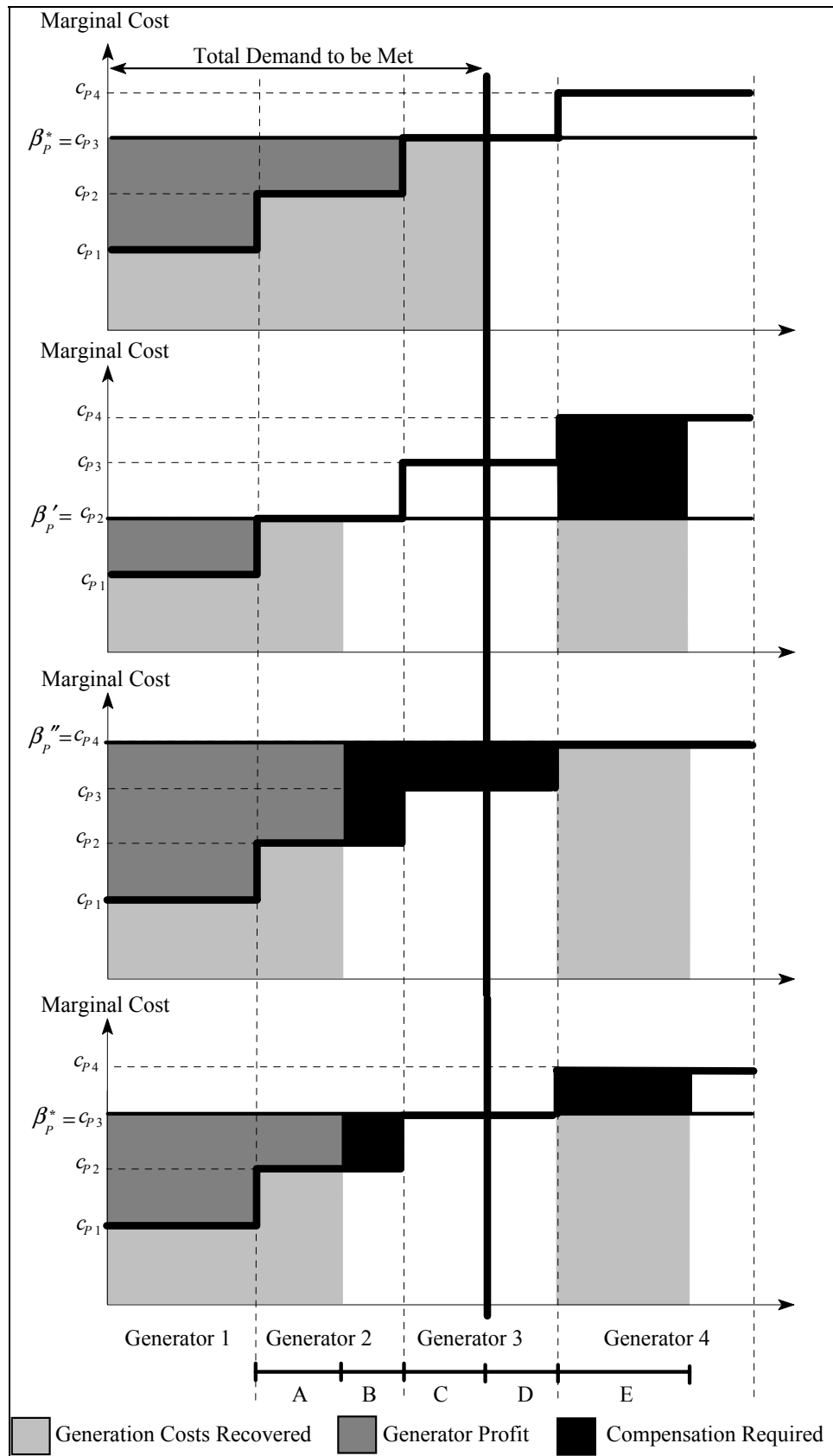
Figure 7.1. Compensation payments for linear power system dispatch.

Figure 7.1 illustrates the process of minimising compensation payments for a simple linear power system example. The top graph describes an optimal dispatch. The specified demand level is met using the cheapest generating units, fully loading generating units 1 and 2 while supplying the marginal unit with generating unit 3. The optimal price, β_P^*, is, therefore, just the marginal cost of operating generating unit 3. The other three graphs depict possible options for pricing for a sub-optimal dispatch. This dispatch is sub-optimal because generating unit 4 has been partially loaded instead of generating unit 3, while generating unit 2 is operating with reduced output. While demand is still met, the cost of the dispatch is now greater than for the optimal case. The second graph depicts what happens when the price, β_P', has been set lower than β_P^*, in this case being the marginal cost of generating unit 2. The black areas indicate the total compensation payments that must be made to off-set the losses incurred by generating unit 4 (the marginal rate of compensation, i.e. the violation price, is just $c_{P4} - \beta_P'$). No compensation need be paid to generating unit 3 for not operating, because, at this price, it should not operate. Likewise, generating unit 2 is marginal, making no profit on the margin, and hence does not need to be compensated for its unused capacity.

The third graph in Figure 7.1 illustrates an alternative situation in which the price is set to equal β_P'', which exceeds β_P^*, equalling the marginal cost of generating unit 4. At this price generating units 1 and 2 make ample profit on what they generate, but generating unit 2 will require compensation for that part of its capacity not called upon. Similarly, generating unit 3 will require compensation for the profit it could have earned had it operated. Generating unit 4 requires no compensation as it is marginal. The final graph shows the compensation required when the price is set to β_P^*. Compensation must be paid for the unused capacity of generating unit 2 and for the uneconomical use of generating unit 4.

The best compromise pricing approach works to minimise the compensation payments and hence minimise the black regions of Figure 7.1. The minimum compensation must occur at a price of β_P^*, since, for a linear primal problem, the best compromise pricing solution must, at optimality, produce prices identical to those for an optimal dual solution to the original primal problem. This can be illustrated by noting that a unit increase in price for the second graph (i.e., price = $\beta_P'+1$) would require an increase in compensation to the owners of generating unit 2 equalling B (referring to the scale below the graphs) while allowing compensation to the owners of generating unit 4 to be decreased by E. But if demand is being satisfied we require that E = B + C, so the net decrease in compensation required is C. Thus moving the price towards β_P^* has decreased the compensation required. Table 7.1 shows, for the sub-optimal dispatch, the change in compensation required for a unit increase and decrease to the prices in each of the last

three graphs of Figure 7.1. Figure 7.2. provides a graphical representation of the results in Table 7.1. Since compensation must be minimised at the point where no better solution can be found by varying the price, this table shows that the optimal solution is to set the price equal to that for the optimal dispatch, i.e. β_P^*.

Declared Price	Net Change in Compensation[109] for:	
	A unit increase in price	A unit decrease in price
β_P'	B - E = -C < 0	A + E > 0
β_P''	B + C + D > 0	E - B - C - D = -D < 0
β_P^*	B + C + D - E = D > 0	E - B = C > 0

Table 7.1. Impact of price changes on compensation.
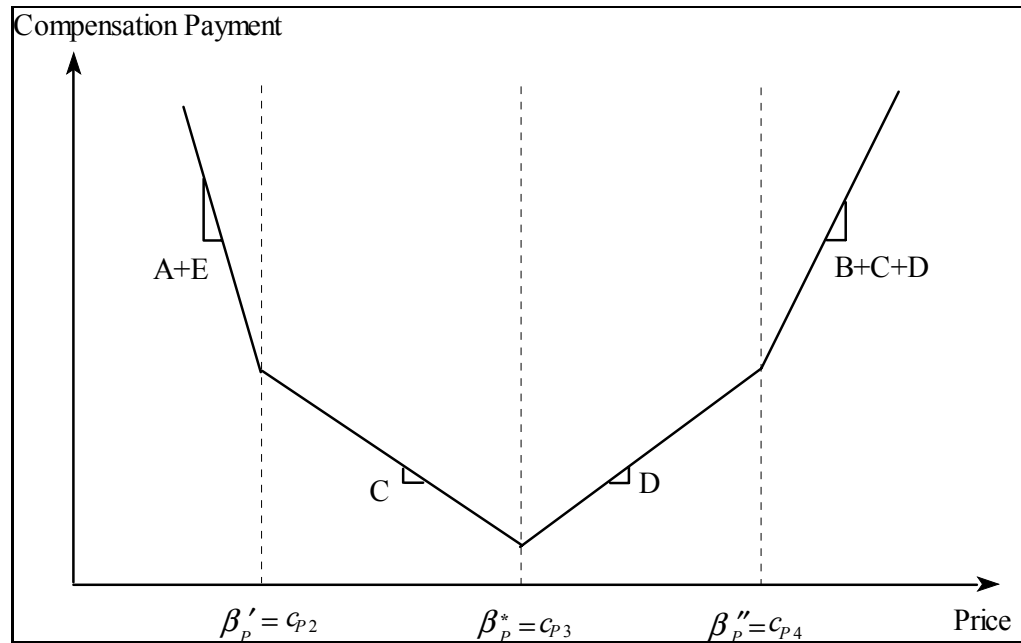
Figure 7.2. Compensation as a function of price.

If the compensation payments suggested in this example were to actually be paid to generating companies, then, regardless of what price is actually used, they will all consider themselves to have received an income consistent with their bids. However, the generating companies will only earn the amount they would have had the dispatch been optimal if the optimal price β_P^* is used as the basis for compensation. This can be seen by noting that

---

[109]   Note that E = B + C.

generating unit 2 earns no marginal profit at a price of β'_P, and a significant profit at a price of β''_P, but should really only earn an intermediate profit at the optimal price of β^*_P.

Figure 7.2. demonstrates that regardless of the method used to dispatch the system, best compromise pricing gives the dispatcher an incentive to set the prices at the point which minimises the compensation required.

7.3.3. Degenerate Problems: A Second Phase

If the optimal dual solution is degenerate, then the solution to the best compromise pricing approach will also be degenerate. Degeneracy will typically arise because a spot price is free to vary between the marginal costs of two generation cost ranges[110]. If there is freedom to vary the spot prices there will also be freedom to vary the compensation payments[111] to each market participant. This introduces the potential for a second phase to our solution process, using an objective function designed to achieve a preferred distribution of benefit between the participants.

This situation is analogous to the standard *ex post* pricing situation where we introduce an arbitrary dual objective function to choose between degenerate dual solutions. In that situation we imposed the complementary slackness conditions directly on the dual constraints, using the "< >" function. Again we can introduce an arbitrary objective function, but rather than using the complementary slackness conditions we can re-solve the dual problem, allowing the prices to vary subject to requiring that the goal programming objective function of (7.22) must always equal the optimal value found by solving (7.22) to (7.24), i.e., the compensation that needs to be injected into the market. If the optimal value of (7.22) is found to be Ω (i.e., this is the minimum compensation required), then the phase two problem is:

---

[110]   This situation is likely to arise when the primal objective function is piece-wise linear.

[111]   Though the total compensation must remain the same in each situation.

$$\mathbf{Optimise} \quad \text{Arbitrary Objective Function} \qquad (7.26)$$
$$\scriptstyle v_x^-,v_x^+\geq 0,\varsigma$$

subject to:

$$\beta_{x\ell} = \sum_{h\in CS} a_{h\ell}\varsigma_h \qquad\qquad \forall \ell \in NDI \quad (7.27)$$

$$\beta_{x\ell} - v_{x\ell}^+ + v_{x\ell}^- = c_{x\ell} \qquad\qquad \forall \ell \in NDI \quad (7.28)$$

$$\sum_{\ell\in NDI}\left(v_{x\ell}^+\left(x_\ell^{max} - x_\ell^o\right) + v_{x\ell}^-\left(x_\ell^o - x_\ell^{min}\right)\right) = \Omega \qquad\qquad (7.29)$$

The prices can now vary as need be to allow the objective function to find the best value of (7.26) with (7.29) ensuring that these prices are consistent with the compensation implied by (7.22) to (7.24).

In effect (7.29) plays the same role as the "< >" functions in the model of Chapter 4. There we only needed to consider the binding primal constraints because these are the only terms that drove the prices. Best compromise prices, however, are driven by constraints which should be binding at primal optimality, rather than by the constraints which are binding at the observed solution. This is why we cannot use the "< >" function of Chapter 4, as all primal constraint terms must be included for best compromise pricing to reproduce the optimal prices.

It may appear that we might as well simply re-solve the original primal problem and determine the optimal prices, rather than using a specialised best compromise pricing version of *ex post* pricing. This might be true for simple linear problems in which all resources are treated in an even handed way for pricing purposes, but we show below that this method can be applied to a wide range of significantly more complex problems.

The "arbitrary objective function" in (7.26) could be the same as that used for conventional dispatch based pricing, such as minimising the network's rent. However, since some compensation payments may accrue to the network, and since we assume that the network is also the dispatcher, we may want to control the trade-off between compensation payments to the network and the market players, while also controlling the distribution of the normal rent. This may be necessary to prevent the network from maximising its own compensation payments. This issue is discussed further in Section 7.6.

7.3.4   An Example of Best Compromise Pricing

To further analyse best compromise pricing it is useful to consider a more complex example. A simple power system dispatch is illustrated by Figure 7.3.
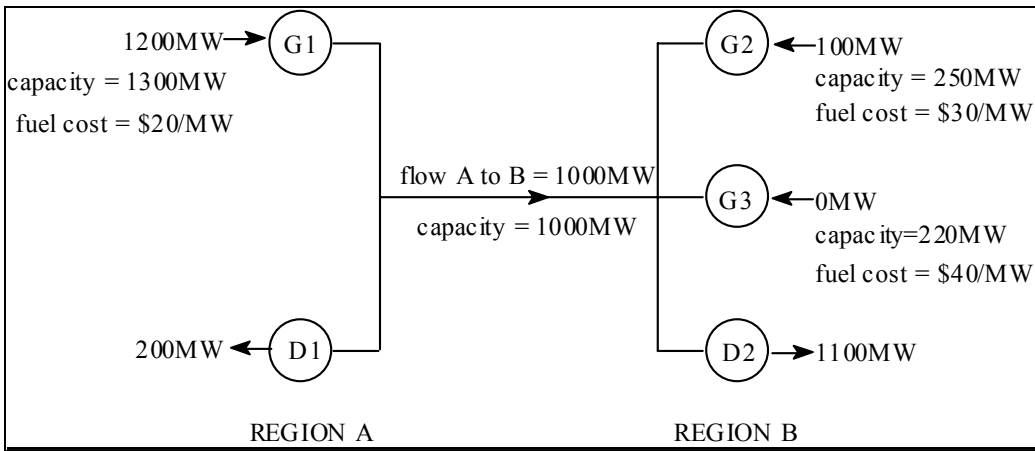
Figure 7.3. An optimal linear power system dispatch.

Region A is primarily a producer of power, while Region B is primarily a consumer of power. We ignore transmission losses, and assume that the consumer's demand levels are inelastic with price[112]. The dispatch shown is optimal, with Region A producing all the power it can, at a cost of $20/MW, given that the amount of power it can export to Region B is limited by the transmission line capacity. The cheapest generation in Region B is used to supplement the imported power so as to meet local demand.

As demand is being satisfied at least cost, marginal cost pricing theory requires that the spot price in Region A equals $20/MW, as at this price the owner of generating unit 1 will have no incentive to increase or decrease generation. In Region B the price must be $30/MW, as at this price the owner of generating unit 2 will have no incentive to change its output, while the owner of generating unit 3 will lack the incentive to begin generating. If we assume a one-to-one relationship between increased generation in Region A and flow on the transmission line then the marginal value of line flow must be $10/MW, reflecting the reduced cost of generation if one more unit of line capacity were to be available. For this example, both operating generating units will just break even, while the grid owner will earn $10/MW $\times$ 1000MW = $10,000.

Now consider the alternative, seemingly inconsistent dispatch shown in Figure 7.4.

---

[112] This assumption has been made for ease of exposition. In general, we could model each consumer as having a demand curve, in much the same way as we represent generation here. Consumers who are misdispatched could be compensated in the same manner as generating companies.
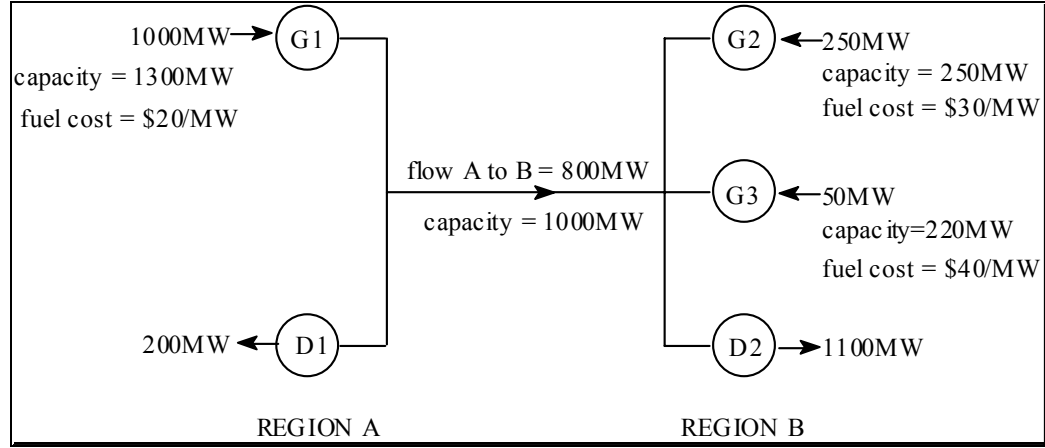
Figure 7.4. An inconsistent observed linear power system dispatch.

The dispatch in Figure 7.4 appears inconsistent because generating unit 1 has the generating capacity to generate, and the transmission line the ability to send, up to an additional 200MW to Region B, allowing a reduction in output by the more expensive generating units there.

Marginal cost pricing theory requires that the price be the same in both regions, as there is no binding transmission constraint. But the fact that generating unit 1 is only partially loaded suggests that this price should equal its marginal cost of $20/MW, while, by the same argument, generating unit 3's marginal cost should set the price equal to $40/MW. This is an example of a pricing inconsistency, requiring the application of best compromise pricing.

If we assume each generating unit and the transmission link to have lower capacity bounds of zero then the best compromise pricing formulation has the form:

$$\underset{v_{\bar{P}}^+, v_{\bar{P}}^-, v_{P}^+, v_{P}^- \geq 0, \lambda_P}{\textbf{Minimise}} \quad v_{\bar{P}}^+\left(\bar{P}^{max} - \bar{P}^o\right) + v_{\bar{P}}^-\left(\bar{P}^o - \bar{P}^{min}\right) + \sum_{i \in PXS}\left(v_{Pi}^+\left(P_i^{max} - P_i^o\right) + v_{Pi}^-\left(P_i^o - P_i^{min}\right)\right) \quad (7.30)$$

Subject to:

$$\beta_{Pi} = \lambda_P \qquad\qquad i \in PXS \cap A \quad (7.31)$$

$$\beta_{Pi} = \lambda_P + \eta_{\bar{P}} \qquad\qquad i \in PXS \cap B \quad (7.32)$$

$$\beta_{Pi} - v_{Pi}^+ + v_{Pi}^- = c_{Pi} \qquad\qquad i \in PXS \quad (7.33)$$

$$\eta_{\bar{P}} - v_{\bar{P}}^+ + v_{\bar{P}}^- = 0 \qquad\qquad (7.34)$$

The sets A and B describe the set of generating units in each region, while $\eta_{\bar{P}}$ is the transmission link multiplier. Simplifying the constraints and substituting the numbers from Figure 7.4 gives:

$$\underset{\substack{v_{\bar{P}}^{+},v_{\bar{P}}^{-},v_{\bar{P}}^{+},v_{\bar{P}}^{-}\geq 0,\lambda_{P}}}{\textbf{Minimise}} \quad 200v_{\bar{P}}^{+} + 800v_{\bar{P}}^{-} + 300v_{P1}^{+} + 1000v_{P1}^{-} + 0v_{P2}^{+} + 250v_{P2}^{-} + 170v_{P3}^{+} + 50v_{P3}^{-} \quad (7.35)$$

Subject to:

$$\lambda_P - v_{P1}^{+} + v_{P1}^{-} = 20 \qquad\qquad (7.36)$$

$$\lambda_P + \eta_{\bar{P}} - v_{P2}^{+} + v_{P2}^{-} = 30 \qquad\qquad (7.37)$$

$$\lambda_P + \eta_{\bar{P}} - v_{P3}^{+} + v_{P3}^{-} = 40 \qquad\qquad (7.38)$$

$$\eta_{\bar{P}} - v_{\bar{P}}^{+} + v_{\bar{P}}^{-} = 0 \qquad\qquad (7.39)$$

The optimal solution has $\lambda_P = \$20/\text{MW}$ while $\eta_{\bar{P}} = \$10/\text{MW}$, i.e. the price in region A is $20/MW and the price in region B is $30/MW. All violation prices equal zero except for $v_{P3}^{-} = \$10/\text{MW}$ (i.e. generating unit 3 should be compensated for the fact that it is generating at a loss)[113] and $v_{\bar{P}}^{+} = \eta_{\bar{P}} = \$10/\text{MW}$ (i.e. the link owner should be compensated for the capacity which should have been used but was not). The optimal objective function value equals $2500, which an inspection reveals to be the difference between the generation costs in Figures 7.3 and 7.4. The resulting prices are, in fact, those corresponding to the optimal dispatch of Figure 7.3.

If the demand in region B had been 1250MW in Figure 7.3 then the optimal dispatch would be the same, except that generating unit 2 would be fully loaded, generating 250MW. In this situation the price in region A would still be $20/MW but the price in region B must be between $30/MW (the marginal cost of generating unit 2) and $40/MW (the marginal cost of generating unit 3), and hence the transmission constraint shadow price can be anywhere between $10/MW and $20/MW. This is a degenerate dual solution. A dual objective function which minimises the network's rent will set $\eta_{\bar{P}} = \$10/\text{MW}$ and the region B price to $30/MW for this optimal dispatch.

If a sub-optimal dispatch occurred, which was, say, identical to that in Figure 7.4 but with a demand of 1250MW in region B and with generating unit 3 producing 200MW, then possible extreme price compensation combinations are shown in Table 7.2.

---

[113] Generating unit 1 should have its output increased, but, as the price in Region A equals its marginal cost, no profit is gained in doing this, so no compensation is required.

β_B $/MW	v^-_{P3} $/MW	$v^-_{\bar{P}}$ $/MW	Rent $1000				Compensation $1000				
			G1	G2	G3	Grid	G1	G2	G3	Grid	Total
30	10	10	0	0	-2	8	0	0	2	2	4
40	0	20	0	2.5	0	16	0	0	0	4	4

Table 7.2. Comparison of rent and compensation

The price discrepancies between the nodes are either explained by a link shadow price of $10/MW, with generating unit 2 being marginal in region B and generating unit 3 being sub-optimally dispatched, or a link shadow price of $20/MW, and generating unit 3 being marginal (though not optimally dispatched, but no compensation is required).

The total compensation payment is $4000 in each case, but the lower region B price solution corresponds to maximising the compensation payout to generating companies ($2000)and minimising the network's rent ($8000 plus $2000 compensation), while the higher region B price solution gives no compensation payouts to generating companies and maximises the network's rent ($16000 plus $4000 compensation). The latter solution seems preferable if we want to minimise the fixed cost payouts for network access, the former is preferable if we want to minimise the network's rent.

7.3.5. Summary of the Theoretical Properties of Best Compromise Pricing for Linear Problems

The following theorem summarises some important properties of best compromise pricing.

Theorem 7.1: Define the best compromise pricing problem to be that of finding a set of market clearing spot prices and (notional) side payments so as to minimise the (notional) side payments which need to be made so that all offers and bids are respected.

Proposition 1: Then for any linear primal problem and any feasible solution to this problem, the optimal spot prices result from the best compromise pricing formulation of (7.22) to (7.24).

Proposition 2: The optimal best compromise pricing objective function must, for a primal cost minimisation problem, equal the difference between the primal objective

function value for the observed feasible solution and the primal objective function value for an optimal primal solution.

Proposition 3: This extra payment must be injected into the market to provide compensation payments for those mis-dispatched relative to the market clearing spot prices.

7.4. Best Compromise Pricing for Piece-Wise Linear Problems

7.4.1. The Impact of Simplified Primal Problem Representations

Piece-wise linear problems are a special case of linear problems, and best compromise pricing could be applied to these as if they were normal linear problems. However it is useful to explore the alternative ways in which best compromise pricing could be applied to such problems, as this provides useful insight into the more general non-linear case.
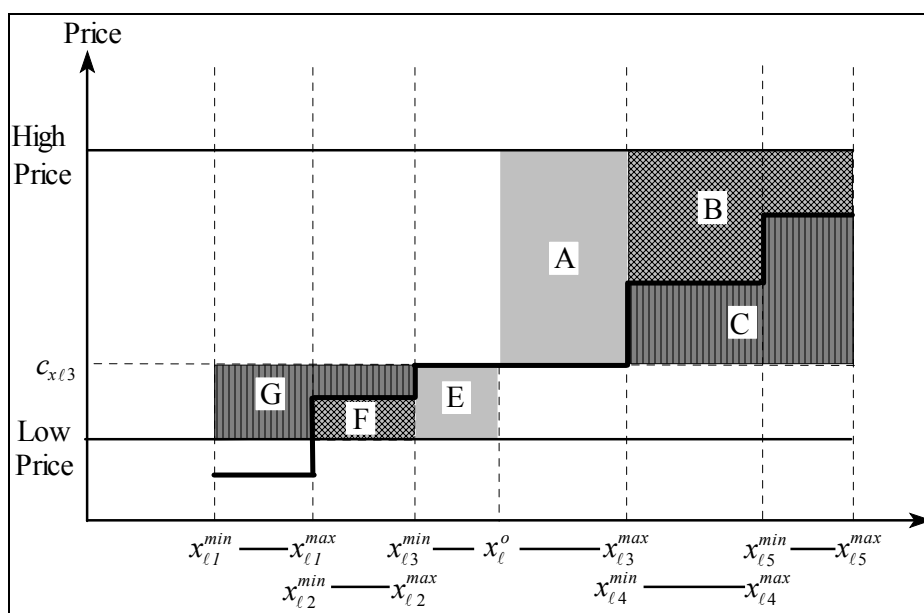
Figure 7.5. Impact of the choice of best compromise pricing bounds on compensation

The bold line in Figure 7.5 depicts a piece-wise linear cost curve for some variable x_ℓ. The range of x_ℓ has been divided into five sections, each section defining one cost range. The observed value of is x_ℓ^o is in the third range, in which the marginal cost is $c_{x\ell 3}$. For this dispatch to be optimal we would require that $\beta_{x\ell} = c_{x\ell 3}$, as we have only partially used the capacity in the third range. If the best compromise price is defined to be the high price shown in Figure 7.5 then there are three different ways of interpreting the compensation that could be paid: These are:

- Full Piece-Wise Linear Representation: This is the ideal approach, whereby each of the cost steps would be treated as corresponding to a different variable, determining the compensation for each range as for a linear problem. This produces the compensation corresponding to the sum of the shaded regions A and B.

- Full Range Linear Approximation: Here we assume that $c_{x\ell3}$ applies over all five cost ranges of x_ℓ, and ignore all bounds except for $x_{\ell 1}^{min}$ and $x_{\ell 5}^{max}$. For the high price case the compensation is the high price less $c_{x\ell3}$ multiplied by $x_{\ell5}^{max} - x_\ell^o$. This corresponds to the combined areas of A, B, and C. For a convex cost function the area A+B+C will always exceed A+B, the optimal compensation. This approach therefore provides more compensation than is strictly required to achieve consistency with the bids and offers.

- Local Range Linear Approximation: This is similar to the Full Range Linear Approximation, but is based on the local region surrounding x_ℓ^o. In effect, we ignore the feasible range below $x_{\ell3}^{min}$ and the feasible range above $x_{\ell3}^{max}$. Now, for the high price case, the compensation is the high price less $c_{x\ell3}$ multiplied by $x_{\ell3}^{max} - x_\ell^o$. This corresponds to the area A. For a convex cost function the area A will always be less than A+B, the optimal compensation. This approach provides no compensation for incorrectly dispatched resources in the feasible range above $x_{\ell3}^{max}$ or below $x_{\ell3}^{min}$, and consequently under compensates market participants relative to their bids and offers .

For a low price situation, similar results are found, with the Full Piece-Wise Linear Representation producing compensation of E+F, while the Full Range Linear Approximation and the Local Range Linear Approximation produce compensations of E+F+G and E, respectively.

It is apparent that the Local Range Linear Approximation places a lower bound on the Full Piece-Wise Linear Representation while the Full Range Linear Approximation places an upper bound on it. Furthermore, since best compromise pricing determines the optimal price for the linear primal assumed, each method will determine the optimal prices corresponding to their respective representations of the primal problem, and, in each case, the compensation will be minimised relative to that price. While the Full Piece-Wise Linear Representation is the only way to reproduce the true optimal prices and compensation payments[114], the Full Range Linear Approximation will produce approximated, though commercially consistent, prices and compensation payments which more than account for the revenue foregone at this approximated price, ensuring that no

---

[114]   And hence is the only approach which will reproduce the optimal distribution of wealth.

market participant makes a financial loss on the margin. The Local Range Linear Approximation, by contrast, provides inadequate compensation to achieve commercial consistency with all bids and offers.

We can conclude, therefore, that it is possible to determine commercially "workable", if not necessarily "optimal", best compromise pricing formulations[115] for some problems without fully representing the underlying primal problem, though the required level of compensation payments must exceed the theoretical minimum commercially consistent compensation level (as defined by the Full Piece-Wise Linear Representation)[116]. This is a particularly important result in power systems, where it may not be possible to define the true primal problem in a single formulation. Conversely, while it is possible to define a best compromise pricing formulation which produces compensation payments which are less than the theoretical minimum commercially consistent compensation level[117], these compensation payments will provide inadequate revenue to achieve commercial consistency with the bids and offers.

7.4.2. Summary of the Theoretical Properties of Best Compromise Pricing for Piece-Wise Linear Problems

The following theorem summarises some important properties of best compromise pricing for piece-wise linear problems.

Theorem 7.2: Given any piece-wise linear primal problem we can, for any feasible solution to this problem, reproduce the optimal spot prices using the best compromise pricing formulation of (7.22) to (7.24).

Proposition 1: For a convex primal cost function the compensation determined by this solution is bounded above by the compensation paid if we treat the primal variables as having the same cost as at the observed solution over the entire range of the variable, and is bounded below by the compensation paid if we treat the primal variables as being defined only over the range in which the costs of the observed solution apply.

---

[115]   E.g. the Full Range Linear Approximation.

[116]   Ultimately we could set the market clearing prices to arbitrary levels, and set the violation prices to account for the discrepancies between these prices and the bids and offers.

[117]   E.g. the Local Range Linear Approximation.

Proposition 2: The prices corresponding to these upper and lower compensation bounds will not correspond to the optimal prices, and neither will they reproduce the optimal wealth distribution. They will, however, produce the optimal prices and compensation payments (and hence wealth distribution) for the simplified primal representations they are based on.

Proposition 3: It is possible to determine commercially "workable", if not necessarily "optimal", best compromise pricing formulations without fully representing the underlying primal problem, though the required level of compensation payments must exceed the theoretical minimum commercially consistent compensation level.

7.5.   Best Compromise Pricing for Non-Linear Problems

7.5.1. Ensuring Boundedness

The discussion of the previous section suggests that best compromise pricing can be applied in the context of the OPF based pricing model of Chapter 4. Provided we can determine a bounded linearisation about an observed solution we can determine best compromise prices. The boundedness issue is important, because, as mentioned above, the pricing model of Chapter 4 is based on a linearisation of the non-linear primal problem about an observed solution, and loses information about more distant solutions in the process. If the observed solution is not optimal, then, by ignoring non-binding primal constraints, and assuming all functions to be linear away from the observed solution, we may lack the information needed to determine prices corresponding to the true optimal solution. This situation is illustrated by Figures 7.6 to 7.8. Figure 7.6 depicts the optimal solution to a non-linear primal problem, the dotted lines describing the contours of the non-linear objective function. The optimum solution lies at the point in the feasible region corresponding to the highest contour of the welfare function. Figure 7.7 describes the linearised primal problem corresponding to Figure 7.6, with only the binding non-linear constraints considered. It can be seen that the optimal solution, and the prices, for this linearisation are the same as for the original non-linear problem[118]. Figure 7.8 depicts the

---

[118]   If there were only one binding primal constraint then the optimal solution would occur where that constraint touched the highest contour of the objective function. It follows that the gradient of the constraint at that point must exactly match the gradient of the objective function. Hence the linearised objective function will exactly coincide with the linearised constraint. Thus the optimal solution to the linearised problem could occur at any point along the linearised constraint. While it may appear that we would then have to constrain the linearised primal solution to equal the original non-linear solution, this is not necessary. The reason for this is that a single dual solution exists at all these possible primal solutions. As we are only interested in the dual solution, and will not be performing any sensitivity analysis, we do not need to introduce any additional primal constraints.

linearisation of a sub-optimal solution to the non-linear problem. Note that the objective function is unbounded, as we have not included any information about the non-binding non-linear constraints.
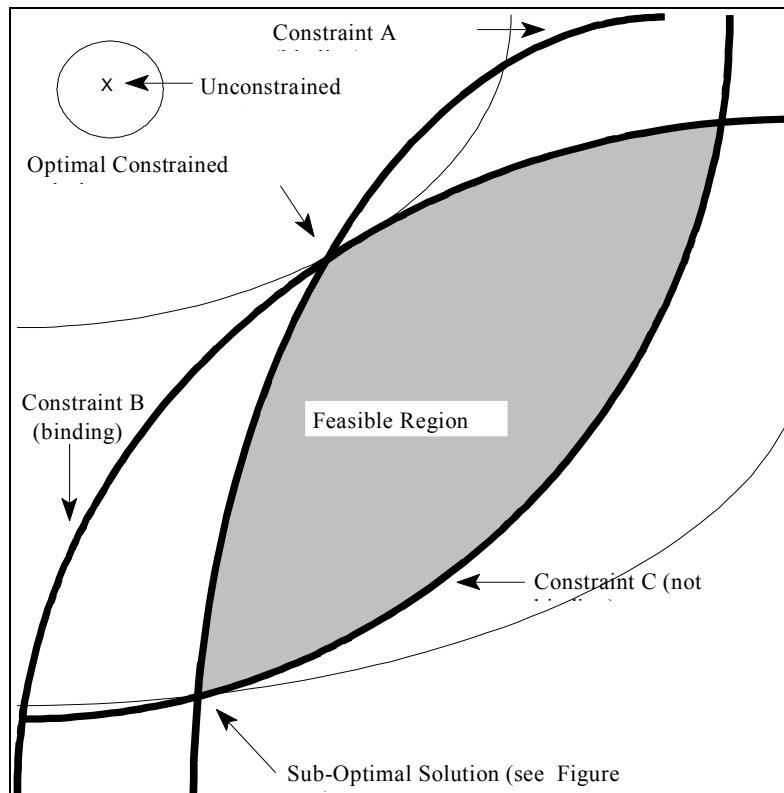
Figure 7.6: A representation of a non-linear primal problem
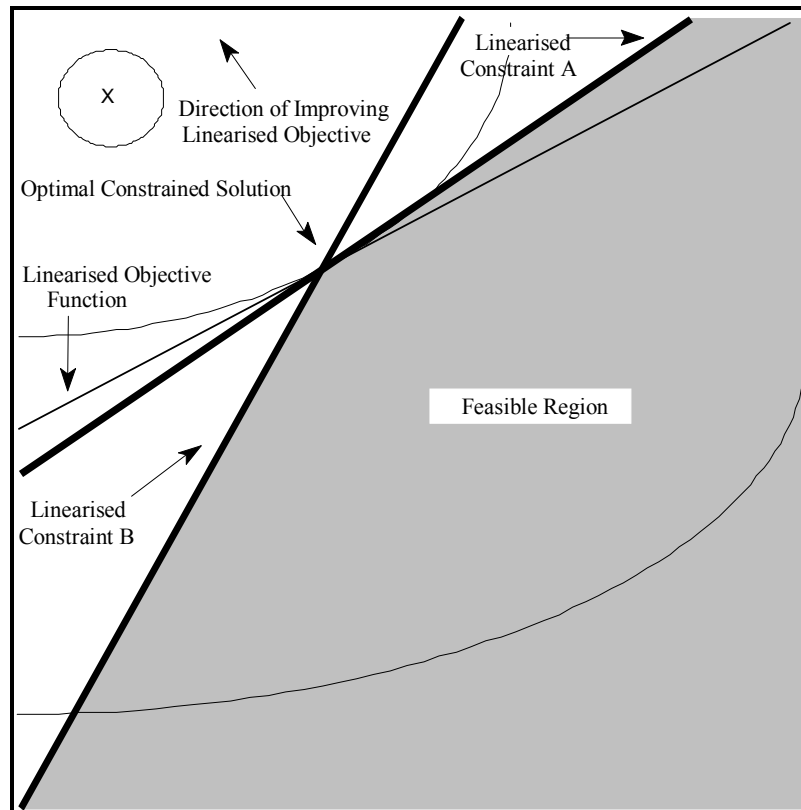
Figure 7.7: The linearisation about the optimal solution of Figure 7.6.
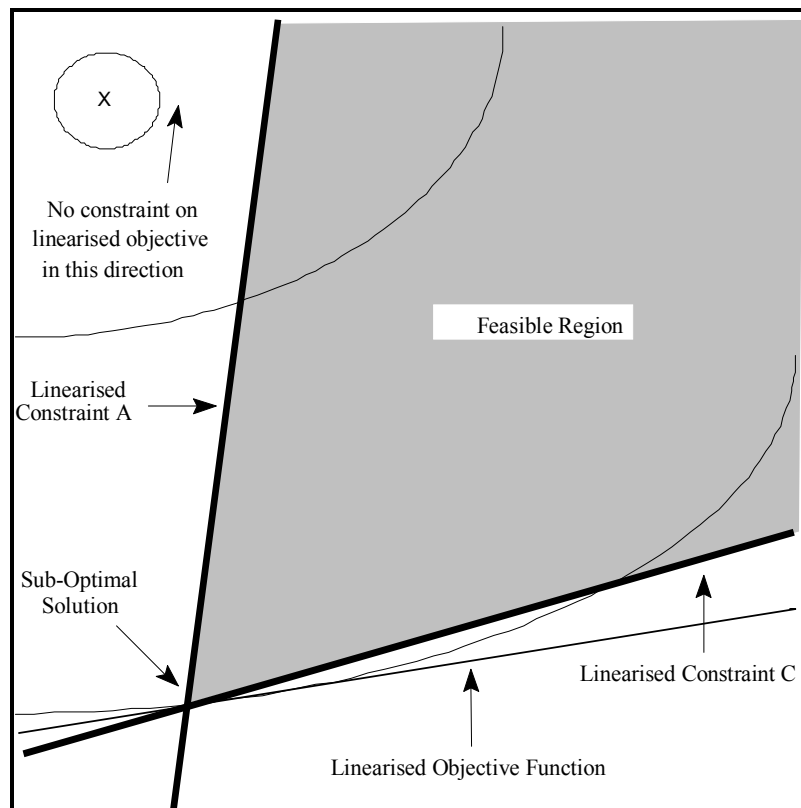
Figure 7.8: A linearisation about the sub-optimal solution of Figure 7.6.

Thus, for sub-optimal non-linear problems, it is crucial to ensure that the linearised primal feasible region is bounded. For a power system boundedness should be provided by the linear bounds on generating unit capacities. Accurately approximating additional primal constraints will enhance the best compromise pricing solution, though, as the approximated primal feasible region will more closely match the true primal feasible region. If implemented correctly, the inclusion of non-binding primal constraints should not alter the prices when the dispatch is truly optimal.

By Theorem 7.2 the compensation payments for such an improvised feasible region may differ significantly from the prices corresponding to the "true" optimum, and some market participants may be over or under compensated relative to that ideal, depending on how the bounds are set. However, the violation prices will still account for the revenue foregone at the approximated prices and given the approximated primal constraints, and will therefore provide enough compensation to ensure that no market participant makes a financial loss on the margin relative to this simplified representation. The crucial issue, though, is whether the market participants are satisfied with the way in which the constraints and bounds on their resources are approximated, as this makes a big difference to their expected returns (see Figure 7.5). The market participants should be satisfied, though, if the approximated bounds force the best compromise prices to be (more than) consistent with their bids and offers, as for the Full Range Linear Approximation. The market participants and dispatcher may need to resolve this issue by negotiation. Assuming this issue can be resolved satisfactorily, and since piece-wise linearisation allows an arbitrarily good approximation of the primal constraints to be made, all parties should be able to accept the compromise prices as the best available prices, particularly if compensation payments are made to those dispatched in a manner inconsistent with these prices.

7.5.3. Summary of the Theoretical Properties of Best Compromise Pricing for Non-Linear Problems

The results of Theorem 7.2. carry over to general non-linear problems, provided that if the "true" primal problem is bounded then the primal assumed for best compromise pricing is also bounded. A major issue in practice, though, is whether the bounds on the approximated primal problem assumed for best compromise pricing purposes will result in compensation payments which are consistent with the actual bids and offers made by the market participants. This can always be achieved (by Theorem 7.2, Proposition 3) if more compensation is paid out than is strictly necessary.

7.6.   Second Order Incentives

7.6.1. Bid and Offer Distortion

It may be suggested that generating companies will manipulate their potential compensation payments by altering their declared marginal cost. For instance, if a generating company expected a price of $50/MW, and had a unit which could produce at a marginal cost of $30/MW, then it would have an incentive to declare a marginal cost for this unit lower than $30/MW, at $10/MW say, since it believes it should be dispatched. If the dispatcher fails to dispatch this unit correctly, and if the expected price transpires, then the company will receive $40/MW compensation, rather than the "natural" $20/MW, on each unit it could have produced. On the other hand, there will be some probability that the generating unit would be dispatched at a price as low as $20/MW, earning no compensation while making a loss of $10/MW. In any case, this "gaming", if judged correctly, does not distort the dispatch and strengthens incentives for good dispatch[119].

7.6.2. Incentives of the Dispatcher

A dispatcher who is also the grid owner may face second order incentives. For instance, if a transmission line is only partially loaded, but the optimal dispatch would have it fully loaded, then the prices in the sending and receiving regions will be inconsistent relative to one another. This is precisely the situation in Figure 7.4. To avoid paying compensation to other parties the dispatcher might claim that, due, say, to maintenance, the line was constrained, with the constraint shadow price accounting fully for the price discrepancy. This is an undesirable situation, however, as the network earns rent which it should not have done, and while the generating companies and consumers are reimbursed in a manner consistent with a constrained link, these financial transactions do not correspond to the correct primal dispatch or pricing solution, disadvantaging some players.

While detailed auditing of the pricing process may limit such price manipulations, the complexity of the power system and the uncertainties involved in dispatching may still leave scope for manipulation.

---

[119]   Overall, the incentives provided by the compensation regime would seem to moderate those from the market.

It might seem appropriate to solve the best compromise pricing model as is, but to not actually pay the resulting transmission compensation to the grid owner. But then the dispatcher ends up with the money, so nothing is achieved if the dispatcher also owns the network. Another possible solution might involve removing the terms relating to transmission lines from the best compromise pricing model objective function, while leaving the constraints - or a representative sample of them - unchanged. This fails to address the issue, though, because now we are minimising the compensation to all parties other than the dispatcher, which is precisely the second order effect we wish to avoid. In particular, the transmission violation prices in the constraint set will be treated as congestion rentals (implying that the lines are constrained), and used as much as possible to explain price variations rather than using the violation prices corresponding to the capacity owned by the market participants.

It might seem, therefore, that we must remove all violation price terms and bounds relating to transmission from the best compromise pricing formulation. This effectively implies that all lines are uncongested, producing no line rentals, even when the line flows are truly constrained. While this result may appear inconsistent with the aim of dispatch based pricing, this is not necessarily the case. For example, to simplify its pricing regime the grid owner may offer market participants the option to pay a fixed fee (or compensation levy) for network usage, and then charge them as if the network is unconstrained. By setting the transmission violation prices to zero and setting the lower and upper transmission flow bounds to negative and positive infinity, respectively, the best compromise pricing problem will determine the market clearing prices consistent with a notional unconstrained grid, and the compensation payments required to resolve any inconsistency between these market clearing prices and the offers and bids. This approach prevents the dispatcher from claiming that a sub-optimal dispatch is constrained.

It is not necessary to resort to a notional unconstrained grid, though, if we recognize that second order incentives cannot be addressed within the standard best compromise pricing model, as it is a "first order" approach. Indeed, all violation prices and bounds must be modelled to reproduce consistent prices. While this might suggest that we must place the role of dispatching the system in the hands of a party which plays no role in the market, an equivalent solution, and one which we explore further, is to use capacity rights (Read and Sell, 1989, Hogan, 1992).

If the grid owner/dispatcher issues capacity rights on the entire capacity of the transmission line in question, and all are sold, then, given a sub-optimal dispatch, the dispatcher has the choice:

- Declare the dispatch to be sub-optimal, and pay compensation; or

- Declare the line to be constrained, collecting the rent on the actual flow, while paying it out to the right holders on the entire capacity.

For example, if the dispatcher owns the transmission line in Figure 7.4 it would be profitable, in the absence of capacity rights, to claim that the line has a capacity of only 800MW, setting the price in region B to $40/MW, and earn a rent of $20/MW on that capacity, or $16,000 in total. By claiming that the line has been dispatched sub-optimally, though, the dispatcher would earn $8,000 rent ($\eta_{\bar{P}}$=$10/MW), but would incur an opportunity cost of $2000 on the unused capacity, and would have to make a payment of $500 to the owners of generating unit 3 (to cover its $500 loss from generating 50MW at a price of $30/MW), giving a net return of $7,500. Hence more rent is earned by claiming the line to be constrained.

The situation is different however, if all the 1000MW capacity of the line has been sold as a capacity right. Now the dispatcher collects $16,000 rent on the line by claiming it to be constrained, but must pay $20/MW times 1000MW to the right holder, meaning that the dispatcher loses $4000. This is more than the $2500 ($2000 on unused capacity rights, $500 to the owners of generating unit 3) that the dispatcher would have to pay by admitting the dispatch to be sub-optimal.

It follows that a dispatcher who has issued capacity rights will attempt to simultaneously minimise the compensation payments and the capacity right payout on capacity in excess of the claimed maximum line flow. Mathematically, this example is consistent with replacing the term:

$$\sum_{\ell \in NDI} v_{x\ell}^{+} \left(x_{\ell}^{max} - x_{\ell}^{o} \right)$$

in (7.22), which we here assume to relate to transmission line upper bounds, with (7.40).

$$\sum_{\ell \in NDI} v_{x\ell}^{+} \left(\left(CAP_{\ell} - x_{\ell}^{o} \right) + \left(x_{\ell}^{\text{capacity right}} - CAP_{\ell} \right) \right) \tag{7.40}$$

That is, the dispatcher wishes to determine values not just for $v_{x\ell}^{+}$, but also for x_{ℓ}^{max}, which the dispatcher surreptitiously treats as a variable, which we call CAP_{ℓ}, so as to minimise the sum of the compensation payments on the sub-optimal line flows (the first term) and the capacity right payouts when capacity rights exceed the value of CAP_{ℓ}[120]. The term $x_{\ell}^{\text{capacity right}}$ is the total capacity right for the line corresponding to variable ℓ. If we assume that $\left(CAP_{\ell} - x_{\ell}^{o} \right)$ must always be non-negative then the dispatcher will always

---

[120] We assume that capacity rights are not a function of the, possibly distorted, line capacities claimed by the dispatcher.

want $\left(x_\ell^{\text{capacity right}} - CAP_\ell\right)$ to be negative, as this will decrease the total payout. Hence the dispatcher will never set CAP_ℓ at a level less than $x_\ell^{\text{capacity right}}$. Also, CAP_ℓ cannot exceed x_ℓ^{max}, i.e.:

$$x_\ell^{\text{capacity right}} \leq CAP_\ell \leq x_\ell^{max} \tag{7.41}$$

The CAP_ℓ terms cancel from (7.40) to give (7.42).

$$\sum_{\ell \in NDI} v_{x\ell}^+ \left(x_\ell^{\text{capacity right}} - x_\ell^o\right) \tag{7.42}$$

Thus, while a dispatcher might be able to manipulate the apparent capacities of transmission lines, capacity right holders are protected from this, being compensated as if the total capacity right is the line capacity. Also the dispatcher will always have incentives to honour those capacity rights, treating the line as if it has at least that capacity.

7.6.3 Industry Capacity Rights

While capacity rights address second order incentives if fully sold, a complication arises if some are not sold, as the dispatcher still faces second order incentives for that capacity from which it earns a rent.

A potential remedy to this problem is to introduce the concept of an Industry Capacity Right (ICR). These rights are defined to be all unsold conventional capacity rights, and would be administered by the industry, excluding the network owner. Since no individual market player has purchased these excess rights we might assume either that some of the market players do not want to deal with the complexity of determining optimal rights, preferring a more conventional share-holding instead, or that these rights correspond to the capacity requirements of the system as a whole which cannot be attributed to the requirements of any individual user (i.e., they are an artefact of the non-linearity).

The rents from the ICRs could be dispersed among the owners in proportion to their share holdings. Since any extra rents due to price manipulations by the dispatcher go to the ICR owners then the incentive to manipulate prices is removed. If the rents on ICRs were to be negative, due, say, to non-linear effects, then the ICR owners would have to pay the dispatcher. The ICR owners will want to make sure that the dispatcher does not manipulate the prices so as to try and unfairly maximise the magnitude of this rent. To avoid this the ICR owners could determine the expected annual value of a negative valued ICR, based purely on first order incentives, and pay this amount to the dispatcher. No further payment is made to the dispatcher when constraints are binding, giving the

dispatcher an incentive to ensure that the magnitude of the actual negative rents do not exceed the amount payed by the ICR owners.

In effect, there is no advantage for the dispatcher in manipulating the rents. The ICR holders can expect the actual rents to exactly match the expect rent if uncertainty is modelled perfectly and the dispatcher obeys only the first order incentives. In the same manner as for conventional capacity rights, any additional (positive or negative) rents received by the ICR owners as a result of price manipulations might be expected to at least correspond to the cost of compensation payments avoided by the dispatcher. Given the potential arbitrariness of compensation payments for some best compromise pricing implementations for non-linear problems it is not clear whether this last result will always be true.

7.7. Conclusions

Past approaches to pricing for economic inconsistencies in dispatches have taken little account of the economic costs of these inconsistencies, instead relying on arbitrary remedies. In this chapter, we have presented a best compromise pricing approach which accounts for the economic costs of dispatch inconsistencies in a way which ensures that prices can be made consistent with the bids and offers of market participants.

Best compromise pricing is based on a goal programming formulation which minimises the (possibly notional) payment that need be made to recover the extra costs arising from genuinely sub-optimal dispatches, or dispatches which appear sub-optimal due to inconsistencies between the assumptions/formulation of dispatch based pricing and reality. Applying this goal programming approach to a truly linear primal problem would give the correct optimal prices and the correct compensation payments. While the approximations involved with non-linear power system models may reduce the accuracy of these results, the goal programming approach can still produce prices which are "workable" in a commercial sense. The prices and compensation will have the following favourable properties:

- All prices will be consistent with the (approximated) power system model.

- No one buying or selling power will make a loss on the transaction that occurs in the sub-optimal solution (relative to the approximated model).

For realistic dispatch problems we may also find that:

- The observed prices will not necessarily match the optimal prices.

- Not all of the welfare loss will necessarily be recovered by compensation.

- The required compensation level may exceed the theoretical minimum level due to approximations in modelling the primal.

If a dispatcher is required to pay compensation, then this might be collected via some annual fee. In minimising the payouts from this fund, an incentive is created for the dispatcher to efficiently operate the system. In situations where the dispatcher is the grid owner there is potential for the dispatcher to make a trade-off between claiming that a dispatch is sub-optimal and claiming that it is more tightly constrained than is really the case. This situation can lead to the dispatcher making excessive revenues at the expensive of generating companies and consumers. Solutions to this problem might include forcing the dispatcher to price according to a notional unconstrained network or to require that the dispatcher does not own the network (or any other capacity). As an alternative solution, we have proposed that all the capacity rights on a line be held by either individual network users, or by a grouping of network users. Market participants have an incentive to acquire such capacity rights even when dispatches are always optimal. Capacity rights effectively transfer excess rents from the dispatcher back to the market participants. As a consequence of these capacity rights the dispatcher's profits are maximised by admitting that the dispatch is sub-optimal in those situations where it truly is.

Generating companies and consumers may have incentives to manipulate their bids and offers so as to get greater compensation than they strictly need if they are mis-dispatched. While this issue requires further analysis, it is apparent that for such manipulation to be worthwhile an accurate expectation of the optimal dispatch is required, and the manipulations must be done in a way which should have the manipulator dispatched according to that expectation, thus effectively improving market efficiency.

Chapter 8

Extensions to Spot Pricing Theory:
Inter-Temporal Constraints

8.1. Introduction

Thus far we have assumed that the dispatch optimisation is based purely on a (possibly notional) OPF, with all costs and constraints explicitly represented, and with the dispatcher able to change the dispatch variables in infinitesimal steps, incurring a cost proportional to that change. As a result we have been able to derive optimal marginal cost based prices. In applying this assumption, we have ignored a number of important aspects of power system operation, including[121]:

- Inter-temporal constraints

  These jointly constrain OPF variables in two or more dispatch periods. In many situations the value of a variable in one period becomes a bound on the equivalent variable in another period. Precisely modelling these constraints may create a very large optimisation problem.

- Stochastic constraints

  These are like inter-temporal constraints but jointly constrain variables in two or more different equilibrium states of the power system. These different states may arise stochastically during the dispatch period. Modelling these constraints precisely may also create a large optimisation problem.

---

[121] In addition to this list we have ignored variations of the dispatch within the dispatch period, treating the dispatch as a "snapshot". This issue is of lesser importance than those discussed here and is referred to in Chapter 4. We have also ignored the impact of harmonics, as, in New Zealand, these are a more significant issue in local distribution networks than in the national grid. Finally, we assume the system is always in equilibrium, ignoring transient periods lasting of an order of a second between equilibrium states.

- Integer dispatch variables

  These variables cannot be varied in infinitesimal steps, and hence are inconsistent with an analysis based on incremental cost variations.

Inter-temporal and inter-state constraints can, in theory, be handled within the framework of an OPF, as discussed below. The inclusion of these constraints is limited in practice partly due to the computational burden they impose, but also due to uncertainty about the nature of possible future states (particularly for inter-temporal constraints) and a lack of reliable data upon which to model the system (particularly for stochastic constraints). Integer variables can not be accurately represented in a conventional OPF, and more importantly, we cannot form a dual to an OPF containing integer variables.

The dispatch based pricing philosophy of Chapter 3 states that the dispatcher must announce prices which provide a consistent economic explanation of the observed dispatch. Thus we suggest that the dispatcher should model the phenomena described above in the most economically consistent manner possible within the theoretical and technical limitations imposed. The best compromise pricing approach of Chapter 7 should be used to account for any remaining discrepancies.

In this and the following two chapters we discuss the nature of inter-temporal constraints, stochastic constraints and integer variables, in turn. We do not attempt to describe a comprehensive framework for addressing these issues, but in each case we suggest, using illustrative examples, how the dispatch based pricing philosophy might be applied. We consider the ideas described in these chapters to be promising avenues for further research. We begin in this chapter by considering inter-temporal constraints.

Bacher (1992) observes that the complexity and high dimensionality of power systems has resulted in the dispatch process being represented by a series of hierarchical models, each valid over a different time frame. Variable outputs of higher level models become fixed inputs for lower level models. Bacher observes that while this decoupling results in some loss of mathematical optimality, practical experience indicates that the results are probably near the global optimum. Bacher defines two basic scheduling categories. The first category deals with time frames of the order of 1 hour to 10 years and uses simplified operational constraints. This time frame includes the unit commitment problem (Baldick, 1995), which determines which generating units are to be committed to the dispatch, but not their actual outputs. The second category involves time frames of the order of 5 minutes to 1 hour with the dynamics of the power system being modelled in detail.

The OPF problem is included in the shorter time frame. The OPF is solved to determine the load flow for a given point in time, but its formulation is based on a solution to the longer term unit commitment problem. It is, therefore, this longer term problem which must be taken into account when modelling inter-temporal effects. In particular, it is necessary to account for the fact that some generating units may need to operate unprofitably in one period so that they can be in position to generate profitably in later periods.

Some of the key inter-temporal links which may exist in a power system are:

- Start up and shut down constraints which require that a minimum time must elapse after a generating unit is committed to run before it can produce any power or before it can be restarted after it is shut down .

- Ramping constraints which limit the rate of change of generation for each generating unit.

- Energy constrained dispatch. This is where a set amount of energy must be produced, or a set amount of fuel consumed, within a given interval, but not necessarily at a constant rate. Thus the dispatcher should vary the amount produced across time so as to maximise benefit.

- Water storage constraints. These constraints limit the storage of water in a dam. The optimal release of water in one period must take account of expected inflows and expected values of power in future periods.

- Hydrological constraints in hydro power river chains. These constraints link the release of water from one dam with the arrival of that water at another dam at a later time. These constraints effectively represent the flow of fuel to the generating units.

In the following section we review the literature on inter-temporal constraints. To provide a reference for later discussion we subsequently present both a ramping constraint problem and an energy constrained dispatch problem, determining the economically optimal solution in each case. We then consider a number of alternative approaches that might be used to account for these constraints in the dispatch process.

8.2.   Review of Pricing Literature on Inter-Temporal Constraints

Caramanis, Bohn, and Schweppe (1982), in the original exposition of the MIT model, presented a multi-period pricing model. This model was based on a primal problem which optimised the system over some time horizon given a set of generic

exogenous random variables which might represent economic factors, weather, and so forth. They derive the optimal spot prices for this problem, these being analogous to those we derive, but are based on the expected impact of marginal changes in the primal variables rather than a deterministic impact. The key impact of inter-temporal constraints is that the value of altering a primal variable value is not just its value in that period, but in all future periods[122]. Caramanis et al. presented this work to highlight the potential of spot pricing to provide good economic incentives, while acknowledging that implementing this model is not trivial. Indeed, Littlechild (1991) argues that this formulation suffers from the so called "curse of dimensionality", the random variates causing the problem size to explode the further into the future the model is projected.

Due to the complexity involved, therefore, inter-temporal constraints are ignored in Trans Power's pricing model implementation, as well as in the implementation of Hogan (1991, 1992) and the MIT model based WRATES program of Caramanis, Roukos, and Schweppe (1989). Hogan, however, suggests that the *ex post* approach can implicitly account for the cost of these constraints on the dispatch because the actual dispatch can be observed. The objective function costs must be modified to represent the actual system state, as well as the costs of the optimal future states. Read (1979) describes how hydro river chain water values can be used in this manner. These water values are the shadow price of water storage constraints[123] and are based on target water levels and/or expected future prices. If accurate, these water values provide the dispatcher with the correct incentives to use the water in an efficient manner. This approach is explored further below.

8.3.   Examples of Inter-temporal Constraints

8.3.1.   A Ramping Constraint

We consider a deterministic, two period, single node, problem. Generating unit 1 has a capacity of 120MW and a marginal cost of $5/MW, generating unit 2 has a capacity of 100MW and a marginal cost of $20/MW, and generating unit 3 has a capacity of 60MW

---

[122]   This is discussed further below.

[123]   These should not be confused with the long term water values specified for major storage reservoirs, though, and can not generally be determined from those by simply pro-rating, for instance.

and a marginal cost of \$40/MW. Generating unit 2 is the only unit subject to an inter-temporal constraint, and can ramp up at a rate of 50MW per period[124]. That is:

$$P_{G2}^1 \leq 100 \qquad\qquad (8.1)$$

$$P_{G2}^2 \leq 100 \qquad\qquad (8.2)$$

$$P_{G2}^2 \leq P_{G2}^1 + 50 \qquad\qquad (8.3)$$

Here the superscript refers to the time period in which the term is defined.   Demand is assumed to be 110MW in the first period and 250MW in the second. If no inter-temporal constraint existed for generating unit 2 then the optimal dispatch would be $P_{G1}^1 = 110$MW while $P_{G2}^1 = P_{G3}^1 = 0$ and $P_{G1}^2 = 120$MW, $P_{G2}^2 = 100$MW, and $P_{G3}^2 = 30$MW. The optimal prices are $\beta_P^1 = \$5$/MW and $\beta_P^2 = \$40$/MW.

We now consider the situation where the ramping constraint is imposed. The price in the first period must still be $\beta_P^1 = \$5$/MW because a higher price would cause generation to exceed demand, while demand could not be satisfied at a lower price, even if generating unit *i* were to produce with full output. It follows that generating unit 2 will make a loss of $\beta_P^1 - c_{P2} = -\$15$/MW on each unit generated in period 1.   Generating unit 2 will not operate in either period if $\beta_P^2 < c_{P2} = \$20$/MW. If generating unit 2 is to operate at all in period 1 then this loss must be offset in period 2, requiring $\beta_P^2 - c_{P2} > \$15$/MW, or $\beta_P^2 > \$35$/MW. If $\$20$/MW$\leq \beta_P^2 \leq \$35$/MW then it will only be worthwhile for generating unit 2 to operate in period 2, and not in period 1.

In this case, to meet the demand of 250MW in period 2 at least cost we again require $P_{G1}^2 = 120$MW, $P_{G2}^2 = 100$MW, and $P_{G3}^2 = 30$MW, implying that $\beta_P^2 = \$40$/MW.   The ramping constraint requires that $P_{G2}^1 = 50$MW so $P_{G1}^1 = 60$MW and $P_{G3}^1 = 0$. The owner of generating unit 2 makes a marginal profit of \$20/MW in period 2 which offsets the loss of \$15/MW in period 1.

It is informative to consider the dual constraints corresponding to this ramping constraint. Re-stating (8.1) to (8.3) in canonical form, and introducing multipliers, gives:

---

[124]   Here we ignore the fact that the generating unit would in reality have to ramp up throughout the first period. While the literature on this topic generally makes this assumption, it is important to note that in practice we might have to set the observed generation levels to be the average generation over the dispatch period, and that the manner in which this done may influence the marginal impact of average generation in one period on the feasible average generation in another.

$$-P_{G2}^1 \geq -100 \qquad\qquad : \upsilon_{P2}^{+1} \qquad (8.4)$$

$$-P_{G2}^2 \geq -100 \qquad\qquad : \upsilon_{P2}^{+2} \qquad (8.5)$$

$$-P_{G2}^2 + P_{G2}^1 \geq -50 \qquad\qquad : \tau_{P2} \qquad (8.6)$$

If we assume, for simplicity, that there is no lower bound on generation in either period, then by applying the methodology of Chapter 4, we can conclude that the prices must satisfy the relationships:

$$\beta_P^1 = c_{P2} + \langle \upsilon_{P2}^{+1} \rangle - \langle \tau_{P2} \rangle \qquad (8.7)$$

$$\beta_P^2 = c_{P2} + \langle \upsilon_{P2}^{+2} \rangle + \langle \tau_{P2} \rangle \qquad (8.8)$$

$$\upsilon_{P2}^{+1}, \upsilon_{P2}^{+2}, \tau_{P2} \geq 0 \qquad (8.9)$$

We have $\beta_P^1 = \$5/\text{MW}$, $\beta_P^2 = \$40/\text{MW}$, $c_{P2} = \$20/\text{MW}$, and, if generating unit 2 is not fully loaded in the first period, $\upsilon_{P2}^{+1} = \$0/\text{MW}$. It follows that $\tau_{P2} = \$15/\text{MW}$ and $\upsilon_{P2}^{+2} = \$5/\text{MW}$. The ramping constraint multiplier value reflects the fact that an increase in the right hand side of (8.6) would reduce the amount by which unit 2 could ramp up to only 49MW, and so it would have to generate one more unit in period one, incurring an additional net loss of $15/MW. It follows that each unit of ramping capacity has a value of $15/MW. The value of generation capacity in period 2 equals $5/MW because if one more unit of capacity were available unit 2 could generate 51MW in period 1, and 101MW in period 2, with a net increase in profit of $5/MW. Thus the incentives are clearly to generate as much as possible in period 2, and accept the consequential loss in period 1, rather than to minimise output in period 1. This follows from the fact that the per unit profit in period 2 exceeds the magnitude of the per unit loss in period 1.

The total net profit made by the owner of unit 2 is:

$$\left(\beta_P^1 - c_{P2}\right)P_{G2}^1 + \left(\beta_P^2 - c_{P2}\right)P_{G2}^2 = (5 - 20)50 + (40 - 20)100$$
$$= -750 + 2000$$
$$= \$1250$$

The rents earned on the constraints are:

$$100\,\upsilon_{P2}^{+2} + 50\,\tau_{P2} = 100 \times 5 + 50 \times 15$$
$$= 500 + 750$$
$$= \$1250$$

As usual the optimal prices which should be paid to the owner of unit 2 account for both the direct operational costs and the rents. It is also apparent from (8.7) that τ_{P2} is just the optimal unit transfer of profits from period 2 to compensate for the marginal financial loss in period 1.

Note that if demand in period 2 had been between 120MW and 220MW then generating unit 2 would be the marginal unit in that period, producing between 0MW and 100MW. The period 2 price would be \$20/MW for demand up to 170/MW as $P_{G2}^2 \leq 50\text{MW}$, i.e. generating unit 2 can satisfy demand in period 2 by ramping up from a zero generation level in period 1, and since it produces nothing in period 1 there is no loss to recover. For demand between 170MW and 220MW we have $50\text{MW} < P_{G2}^2 \leq 100\text{MW}$ and the period 2 price must be \$35/MW, i.e. generating unit 2 must produce power in period 1 to allow it to contribute in period 2[125]. In the first case generating unit 2 would just break even overall, while in the second case it just breaks even on the first 50MW of generation but makes a marginal profit of \$15/MW on any additional generation. This additional profit would not be earned if the ramping constraint had not been imposed, as generating unit 2 would not operate in period 1 and would be marginal in period 2, and hence a second order incentive may exist for the owner of generating unit 2 to exaggerate the form of the ramping constraint so as to maximise the additional profit in period 2.

8.3.2. An Energy Constrained Dispatch

We now consider an energy constrained dispatch where the constraint requires that 150MW be generated in total over the two periods by generating unit 2. In this case (8.6) would be replaced by:

$$P_{G2}^1 + P_{G2}^2 = 150 \qquad (8.10)$$

The optimal solution would again have $\beta_P^1 = \$5/\text{MW}$, $\beta_P^2 = \$40/\text{MW}$, $P_{G1}^1 = 50\text{MW}$ and $P_{G2}^2 = 100\text{MW}$. The dual pricing constraints are now given by:

$$\beta_P^1 = c_{P2} + \left\langle v_{P2}^{+1} \right\rangle + \left\langle \xi_{P2} \right\rangle \qquad (8.11)$$

$$\beta_P^2 = c_{P2} + \left\langle v_{P2}^{+2} \right\rangle + \left\langle \xi_{P2} \right\rangle \qquad (8.12)$$

$$v_{P2}^{+1}, v_{P2}^{+2} \geq 0 \qquad (8.13)$$

---

[125]   A situation might arise where generating unit 2 is marginal in both periods. For example, if demand in period 1 had been only 40MW while demand in period 2 had been 200MW then generating unit 2 could produce 40MW in period 1, being the marginal source of power, while in period 2 generating unit 1 could produce 120MW and generating unit 2 produce 90MW. The price in both regions can now take any value which ensures that generating unit 2's marginal profit in period 2 offsets the marginal cost in period 1. At one extreme, if generating unit 1 is to stay out of the dispatch in period 1, the price can be \$5/MW in period 1 (marginal loss=\$15/MW) and \$35/MW in period 2 (marginal profit =\$15/MW), while at the other extreme, if generating unit 3 is to stay out of the dispatch in period 2, the period 1 price can be \$0 (marginal loss=\$20/MW) while the period 2 price can be \$40/MW (marginal profit=\$20/MW).

Where ξ_{P2} is the multiplier on the energy constraint and is unconstrained in sign. Given that $c_{P2}=\$20$/MW and $v_{P2}^{+1}=\$0$/MW (since generating unit 2 is not fully loaded in the first period) (8.11) requires that $\xi_{P2}=-\$15$/MW, reflecting the fact that an increase in the right hand side of (8.10) can only be meet by generating at a loss in period 1. The capacity multiplier in the second period must be $v_{P2}^{+2}=\$35$/MW, as one more unit of capacity would allow production in period 2 to increase (with a profit of $20/MW) while generation in period 1 could decrease (saving $15/MW).

8.4. Offer Price Modification

Rather than explicitly informing the dispatcher that a ramping constraint exists, the company which owns generating unit 2 may instead modify this units offer in a manner which reflects the impact of the constraint. If this company expects $\beta_P^2=\$40$/MW then it will be prepared to make a loss of up to $20/MW in the first period (provided $\beta_P^1 \geq \$0$/MW) if this allows it to take advantage of the profit to be made in period 2. However, the maximum amount upon which it should be prepared to make a loss in period 1 is 50MW, as any generation above this does not increase units 2 generation potential in period 2. Hence the following offer might be issued for generating unit 2[126]:

Period 1[127]

$$c_{P2} = \$0 \, / \, \text{MW} \qquad\qquad \text{for } P_{G2}^1 \leq 50\text{MW}$$

Period 2

$$c_{P2} = \$20 \, / \, \text{MW} \qquad\qquad \text{for } P_{G2}^2 \leq 100\text{MW}$$

If the expected prices eventuate then generating unit 2 is guaranteed to recover the losses it incurs in period 1. A complexity with this approach is that if the expected prices eventuate but the dispatcher determines, incorrectly, that P_{G2}^1 should be 40MW then, if physically dispatched at this level, the ramping constraint would make it impossible for an output of 100MW to be achieved in period 2. In practice, this could be avoided if the generating company is forewarned of the proposed dispatch, and has the opportunity to insist on operating at 50MW in the first period. A more extreme version of this problem

---

[126]   This offer would also satisfy the energy constraint of (8.10).

[127]   The zero bid price in period 1 is purely coincidental. This value is just the expected profit in period 2 less the marginal cost of generating unit 2. If the expected profit were larger then the period 1 offer price would exceed zero.

occurs if the generating company's expectation of the price in period 1 was inaccurate, and the period 1 price was actually less than zero. In this case the company would prefer that unit 2 was not dispatched in either period, but may be forced to insist on being dispatched in period 1 so as to meet its commitments in period 2.

This example is rather simplistic, as we have ignored the initial state of generating unit 2 in period 2, and the ramping constraint between period 0 and period 1, and we have imposed no end conditions. In practice generating unit 2 will have an initial, period 0, generation level, while its owner will have an expectation of how prices will develop over the coming periods (these expectations being a function of how unit 2 is dispatched) and may want unit 2 to be operating at a particular level at the end of the horizon. Given this information the owner of generating unit 2 must determine offer price and generation capacity combinations for the time horizon which reflect the impact of the ramping constraint. While this problem may appear complex, it is similar in form to the problem of determining water values in hydro systems (Read, 1979) and can be solved using dynamic programming techniques for example. We note that the same problem must be solved in a centralised dispatch, but for all generating units simultaneously.

8.5. Quantity Targeting

A quantity targeting approach would be a variant of offer modification which places the primary focus on modifying the offer quantity rather than the offer price. For example, the owner of generating unit 2 could simply determine the quantities it wants to generate and fix its offer quantities at these levels, giving the dispatcher no choice in how it should be dispatched. Thus, given the same price expectations as assumed above, the owner of generating unit 2 could make an offer of $P_{G2}^1 = 50\text{MW}$ and $P_{G2}^2 = 100\text{MW}$. Given such an offer the dispatcher could justifiably assign a shadow price, unconstrained in sign, to generating unit 2's capacity in each period. Applying the conventions of Chapter 4, a positive shadow price corresponds to the marginal profit earned by generating unit 2 in that period, while a negative shadow price describes the marginal loss. Provided that these shadow prices do not appear in the dispatch based pricing objective function, these shadow prices can vary freely so as to explain any difference between the spot prices and the offer prices. In effect, the pricing model will never consider generating unit 2's offer price in determining the spot prices. Ultimately then, by issuing this offer, the owner of generating unit 2 is signalling to the dispatcher that it accepts all the financial risk associated with this offer.

8.6.   Explicit Notification of Ramping Constraint

Another approach to this problem would be for the owner of generating unit 2 to inform the dispatcher, via its offer, of the requirement that it be operated in a manner consistent with its ramping constraint. The dispatcher would then have the responsibility to dispatch the generating unit consistent with this constraint, as well as to determine prices which explain the economics of that dispatch.

Multiplier: (τ_{P2}) ($/MW)	Period 1 $(\beta_P^1 =\$5/MW)$		Period 2 $(\beta_P^1 =\$40/MW)$		Total Compensation Payment Required ($)
	Marginal Profit[128]: $\beta_P^1 + \tau_{P2} - c_{P2}$ ($/MW)	Dispatch Error (MW)	Marginal Profit: $\beta_P^2 - \tau_{P2} - c_{P2}$ ($/MW)	Dispatch Error (MW)	
0	-15	0[129] 50	20	-50 0	1000 750
5	-10	0 50	15	-50 0	750 5000
10	-5	0 50	10	-50 0	500 250
15	0	0 0	5	-50 0	750 0
20	+5	-100 -50	0	0 0	1000 250

Table 8.1: Impact of compensation on the shadow price of a ramping constraint

While a multiple period pricing model would be desirable in this case, the dispatcher could choose to simply estimate the inter-temporal shadow price, and then apply best compromise pricing to determine any compensation that might need to be made.

Table 8.1 demonstrates how a dispatcher might go about determining both the dispatch and the ramping constraint multiplier[130] for our example when the prices are

---

[128]   Where $c_{P2} =\$20/MW$.

[129]   The first entry in each cell relates to a dispatch of 0MW in the first period and 50MW in the second, while the second entry relates to 50MW in the first period and 100MW in the second. The stated values are the observed output for that period less the optimal output.

[130]   An analogous method could be applied if we had an energy constraint instead of a ramping constraint.

given[131]. In Table 8.1 we consider two extreme dispatches, one with 0MW produced in the first period and 50MW in the second, the other with 50MW produced in the first period and 100MW in the second. The dispatcher varies the multiplier value so as to minimise the compensation payment that would need to be made. It is apparent that a dispatch of 50MW in the first period and 100MW in the second, combined with a multiplier value of $15/MW, minimises the compensation required.

It is no coincidence that this solution is optimal. Compensation is minimised for this solution as it is the only solution for which the power prices can cover both the generating costs in both periods as well as the value of the inter-temporal constraint. This example demonstrates how the best compromise pricing approach can produce consistent prices, even for a non-optimal inter-temporal dispatch, while the compensation payments encourage the dispatcher to try and find the optimal dispatch.

While a full multi-period best compromise pricing approach would be the idea way to implement this approach, it does not avoid the complexity of modelling inter-temporal constraints. While abstracting away from this complexity may make it easier for the dispatcher to determine prices and compensation payments, the need to maintain consistency with the bids and offers means that compensation may not be set to its minimum level. The possibility of paying more compensation than is strictly required should give the dispatcher an incentive to maximise the accuracy with which inter-temporal effects are modelled[132].

8.7.   Conclusions

Inter-temporal constraints have a major impact on power systems, but the complexity involved in modelling them has limited the extent to which they have been represented in pricing models. We have presented two approaches which might be applied to inter-temporal problems when a detailed multi-period representation is considered too complex. One approach requires that generating companies modify their offers so as to account for the inter-temporal effect. For example, a generating unit owner faced with ramping

---

[131]   If the prices are not given, which will be the typical case, then the dimensions of Table 8.1 could be increased to allow a range of pricing solutions to be tried. Ideally, a two period best compromise pricing formulation would be used, though this is equivalent to solving a multi-period dispatch problem.

[132]   In practice, the levy placed on the market to collect the revenue required for compensation payments should reflect the accuracy with which the market expects the complexities of power systems to be modelled. The dispatcher could (potentially) be held liable for the additional compensation required if the expected level of modelling accuracy is not achieved.

constraints might offer to operate below this units marginal cost in some periods if this allows this unit to achieve maximum output in profitable periods which follow. While this may lead to some loss in coordination efficiency, making it necessary to check the consistency of the pre-dispatch, the onus is on generating companies to accurately predict the evolution of the dispatch and to behave according to that prediction, thus reinforcing the incentive for efficient operation.

A second approach is for the generating companies to simply dictate to the dispatcher the levels at which they wish to generate in each period. In this situation the generating companies bear all the risk, as the dispatcher can activate a shadow price, unconstrained in sign, on these constraints, effectively removing the relevant generating units from the pricing problem. This is analogous to the treatment of voltages and reactive power injections which have not been optimised, as discussed in Chapters 3 and 4.

A further approach is for generating companies to explicitly inform the dispatcher of the ramping constraint, and make it the dispatcher's responsibility to determine generation levels, and the corresponding prices, over time. This approach leaves the issue of how the dispatcher determines the dispatch open, though the compensation payments which would be implied if best compromise pricing were to be used encourages the dispatcher to find solutions which are as near to optimal as can be found. The ability of best compromise pricing to find the minimum the compensation payment required may be limited, though, by the accuracy with which inter-temporal constraints are modelled.

What ever approach is used, effectively the same judgements must be made, and it is moot whether it is more efficient or effective for these to be made by a central dispatcher or by many independent market participants.

We have shown that imposing a ramping constraint can increase the profit earned from a constrained generating unit, and hence that a second order incentive may exist for generating companies to exaggerate the form of ramping constraint so as to maximise their profits. The offer modification approach may go some way to limiting this effect, because in making such modifications generating companies will consider their actual ramping limitations.   It is not clear whether or not this advantage outweighs the cost of the coordination inefficiencies that this approach may create.

Chapter 9

Extensions to Spot Pricing Theory: Stochasticity

9.1. Introduction

The previous chapter assumed perfect foresight, but power systems are operated under uncertainty, and the dispatcher must prepare for a variety of situations, incurring costs by doing so. Some uncertain aspects of short-run power system operation are:

- future demand,

- equipment failure, and

- hydrological inflows.

In this chapter we discuss the use of dispatch based pricing to account for stochastic aspects of the dispatch. In particular, we focus on the stochastic multiple contingency problem, a problem which, in the time frame of the dispatch, is the most significant, and potentially expensive, source of uncertainty in the New Zealand power system[133].

Our analysis of reserve pricing in Chapter 6 assumed reserve must be held to satisfy a target reserve requirement, while ignoring the optimisation of that requirement and the possibility of that reserve actually being required. In effect we assumed that disruptions would not occur often enough that the costs incurred during disruptions would have any significant effect on the choice of reserve options. By ignoring the possibility of a contingency actually occurring, the reserve sources with the lowest "stand-by" cost will always be the most favoured, and cheap unreliable generating units will more readily be base loaded than more expensive but more reliable generating units. However, actually calling upon such sources may be more expensive than calling on other sources with higher stand-by costs. If we are to correctly trade-off the usage costs and stand-by costs of

---

[133]   The stochastic reserve problem we refer to relates to equipment failure, rather than to the problem of stochastic load forecasting in situations where capacity may be inadequate to meet demand.

various reserve sources, so that the dispatch decision can be made optimally, we must consider a range of possible random disruptions in any given dispatch period, and the available response options.

We explore these issues after reviewing the stochastic reserve problem and the literature on pricing for reserve under uncertainty. The relationship between dispatch based pricing and the stochastic reserve problem is then discussed, with an "Optimal Reserve Targeting" model, based on a model described by Caramanis et al. (1987), being proposed as a reasonable compromise between detail and practicality. The dual of this scheduling model is analysed, and we demonstrate that the common assumption that energy prices rise during contingencies is not necessarily true. Finally, we briefly discuss the implications of the longer term effects stemming from a contingency.

9.2.   Stochastic Reserve Scheduling

The stochastic reserve problem involves determining what mix of reserve sources, including generator spinning reserve, load curtailment, and blackout, each having a different cost structure, is best suited to meeting the range of possible contingencies. It is also necessary to ensure that both pre- and post-contingency power flows are feasible. An added complexity is that sources of reserve can also be eliminated by contingencies, making reserve availability stochastic. This problem also has an analogy with inter-temporal constraints in that we must consider the linkage between the base state and the contingency state. Further, following the re-establishment of equilibrium, it may be necessary to bring more generating units into the dispatch to replace expensive reserve options, such as blackout. It may be several hours before these generating units can reach the required output. This introduces longer term constraints, which may include integer variables, and apparent sub-optimality, into the reserve problem.

The theory of "optimally" meeting system security requirements is reviewed by Stott, Alsac, and Monticelli (1987). They distinguish between two approaches for scheduling reserve capacity, both for generating units and transmission lines. The simplest, and most commonly used, approach is a security constrained dispatch, whereby target security levels are imposed on the dispatch, much as discussed in Chapter 6. A security constrained dispatch does not explicitly model contingent states. A more general approach is a contingency constrained dispatch, which can be modelled using a *Contingency Constrained Optimal Power Flow* (CCOPF) formulation. The CCOPF can be viewed as an OPF formulation constrained to be feasible for a set of possible contingencies. The CCOPF described by Stott et al. is based purely on engineering requirements, and does not consider the probabilities of contingencies occurring, or the costs associated with

operations in the contingency states. While this formulation might introduce millions of extra constraints into the primal OPF formulation, only a few of these are ever likely to be binding at optimality. An iterative solution process is used to solve this problem. Given an initial dispatch, the violated, or near binding, security constraints are identified. A new problem is then solved subject to the pre-contingency, or base state, constraints, and the post-contingency constraints of the selected contingencies. This procedure is repeated until no violations remain.

Stott et al. report that a conservative formulation of the CCOPF has the system dispatched so that the base state has the capability to respond to all contingencies, so no major control action is required when a contingency occurs. For instance, if responses based on frequency drops or initiated by automatic generation controls (AGC) give rise to feasible contingency states, with no system limits violated, then system security is maintained without any direct action being required of the dispatcher. A more risky, though cheaper, approach is to not cover all possible contingencies but rather to re-dispatch the system immediately once a contingency has occurred so as to alleviate any constraint violations.

An accurate economic representation of the stochastic contingency problem is more complicated than the form described by Stott et al., involving the minimisation of the expected costs over all states, rather than just the base state. This may greatly increase the number of contingency constraints that are binding. Such general dispatch models are not used in practice, presumably because of their size and complexity.

9.3. Review of Pricing Models of Stochastic Reserve Dispatch

Caramanis, Bohn, and Schweppe (1987), building on the ideas of Caramanis et al. (1982), describe a stochastic reserve pricing model, explicitly modelling all possible contingent states of the system via linear security constraints. Prior to dispatch, and in addition to normal energy market transactions, each market participant pledges reserve (an increase in generation or a decrease in demand) and sets a minimum reserve price for the use of that reserve when called upon. Given the pledged reserve, a price is determined for reserve made available, this price being the shadow price on the primal constraint requiring that some minimum reserve availability be achieved. During the period, as actual (stochastically distributed) reserve requirements become apparent, reserve is called upon and a price for this power is set in accordance with the original reserve offer prices. Consumers and generating companies maximise their individual expected welfare with respect to their demand levels, and reserve availability as well as to the distribution of possible reserve usage, the expected payment (positive or negative) to a market participant

equalling the value of that participant's net energy supply in the absence of a contingency, plus the value of its marginal contribution to meeting the system's total reserve availability requirements, plus the expected value of the extra generation it actually provided (or required) during the period.

Berger and Schweppe (1989) present a more radical implementation, which uses prices to guide a power system through the short transient phase following a contingency. Information about the willingness of each market participant to react to a particular price change might be provided by a "black box" representation of that participant. Berger and Schweppe acknowledge that this model is only speculative, but argue that further research seems justifiable.

Hogan (1991) proposes an alternative approach, which uses a CCOPF based primal formulation to determine *ex post* prices which account for post-contingency power flow feasibility requirements. Given an observed solution it is possible to identify the security constraints which were binding and to account for them in the prices accordingly. To model generating unit failure Hogan proposes defining the losses relative to the loads prior to the contingency occurring but the transmission constraints relative to the post-contingency power flows. This approach is consistent with a CCOPF formulation which typically only models pre-contingency losses. Hogan does not explicitly model ramping constraints which might be imposed on generating units (though in principle this can be modelled by modifying the offer prices[134]), and makes no explicit account of the contingency probabilities.

The models of Caramanis et al. and Hogan address different aspects of the stochastic contingency problem. On one hand, Caramanis et al. (and Berger and Schweppe) focus on the providers/users of reserve, and present a model which explicitly accounts for stochasticity. On the other hand, Hogan deals more with transmission constraint issues, and his model is based on the requirement of maintaining feasible power flows during contingency, with no regard for probabilities. Hogan's proposal is based closely on present engineering practice while the proposal of Caramanis et al. optimises the operation of the power system. Hogan prices for what actually happened, subject to expectation of contingency requirements, while Caramanis et al. (1987) price for reserve availability prior to the dispatch, and for reserve usage during the dispatch.

---

[134]   As discussed in Section 8.4.

9.4.   Dispatch Based Pricing and the Stochastic Reserve Problem.

The models of the previous section abstracted away from the "ideal" reserve pricing model. Ignoring transient effects, the "ideal" model would require that demand be satisfied in each system state, whether by generation, load curtailment, or blackout, and that the most economical mix of reserve capacities be maintained in that state so as to allow demand to be satisfied most economically if the system state changes due to any additional outage[135]. Unlike the model of Caramanis et al. (1987), this model is recursive, with each state being constrained in an analogous manner to the "no outage", or "base state". If we assume that this recursion continues until a state is reached in which all generating units have failed, then each state will be more tightly constrained, and hence will have a higher energy price, than its predecessor. This "ideal" model, therefore, exhibits the type of pricing behaviour often assumed in loose discussions of reserve issues, namely that the price in a contingency state is higher than if no contingency has occurred, and that the contingency state prices rise with the size of the contingency.

While economically elegant, the complexity of such an "ideal" model is likely to be so great as to make its use impractical, especially when one considers the relatively unambitious goals of contingency models actually used by engineers. We must therefore consider simpler models which represent the basic engineering objectives, which can provide a consistent dispatch based pricing interpretation, and which can be solved in practice. A model of the type proposed by Caramanis et al. (1987) would appear to be a practical compromise between the simplicity of the model of Chapter 7 and the complexity of the "ideal" model. The Caramanis et al. model sets the reserve to be carried in the base state, and requires that any of this capacity which is unused in a contingency state must still be carried as reserve. This latter condition is very important, as it gives the system some ability to respond to further contingencies, even if these are not explicitly modelled[136]. The reserve which is carried in the base state does not need to be an arbitrary choice, as in Chapter 6, but can be optimised so as to minimise the cost of the dispatch[137]. We refer to this type of model as an "Optimal Reserve Targeting", or ORT, model. This

---

[135]   Such a model has been derived in a discussion draft by Ring (1995).

[136]   If this condition were not imposed then the optimal response to the smallest contingency would be to meet it with the cheapest merit order dispatch and to shut down all surplus generating units, leaving the system with no further reserve capabilities (apart from the use of the generating unit which is marginal in this state). Thus the price would almost always fall during contingencies.

[137]   If the base state reserve is insufficient to cover an outage then blackout must be employed.

model requires that the dispatcher ensures that the mix of reserve sources available is economically justifiable given the range of possible contingencies.

A crucial assumption of this and the other models described in the literature review is that the system responds optimally when an outage occurs, at least with respect to the formulations assumed. One interpretation of this could be that it is possible to monitor and control the system perfectly during a contingency, allowing the dispatcher to choose the cheapest response. This may not be the case in reality, though. A drop in system frequency may be the only signal that something has failed, with turbine governors responding automatically. Hence the response may be initiated before the dispatcher knows what has failed, and this creates uncertainty as to whether the response is optimal or not.

An alternative interpretation might be to view the role of the dispatcher as optimising the trigger mechanisms for the automated responses. In effect, the dispatcher must determine the best form of response without anticipating that a specific contingency is actually going to occur. For example, if turbines and curtailable load typically respond automatically to sudden frequency drops, the dispatcher should optimise the frequency trigger point[138] at which each of these sources initiates a response so as to minimise the expected cost of reserve provision over all contingencies, but can not be expected to intervene to provide an "optimal" response to a particular emergency. This approach may produce higher costs than if the dispatcher could optimise the response in real time, but this simply reflects a real constraint on the system, and hence can be justified economically.

If we assume the existence of an accepted industry measure of the recent reliability of generating units, the role of dispatch based pricing can be seen as one of determining prices for energy and reserve which are consistent with this measure and the bids and offers of the market participants. Given this viewpoint, an appropriate pricing model might have the form of that in Chapter 4, but would include CCOPF type constraints on power flows but with ORT type constraints dictating the utilisation of spinning reserve, load curtailment, and blackout. Some variant of the constraints of the reserve model in Chapter 6 might fulfil this latter role, at least in New Zealand. In the primal dispatch problem we would require that the capacity of spinning reserve and load curtailment[139]

---

[138]   If automatic generation controls (AGC) are employed then the response may not be initiated directly by a frequency drop, and rather than setting the frequency trigger points the dispatcher can instead optimise the AGC response.

[139]   This problem may be simplified in practice if load curtailment availability is based on a long term contract, fixing the load curtailment availability in the short run.

scheduled by the dispatcher be sufficient to cover the minimum frequency and surplus reserve constraints corresponding to a range of outages based on different combinations of generating unit and transmission failures, with blackout being the option of last resort[140]. The probability weighted cost of calling on the reserve sources would be reflected in the primal objective function, and hence the price bounds on the dual.

Finally, the dispatch based pricing philosophy provides two ways of relating the prices to the dispatch. Prices can be:

- based on dispatch <u>outcomes</u>, so that those generating companies which choose, or are required, to provide reserve would receive no reward unless a contingency actually occurred, in which case prices would have to rise to very high levels during this contingency. The price for reserve actually called on for each contingency would equal the marginal cost of matching an increase in the contingency size.

- based on dispatch <u>intentions</u>, so that those generating companies which choose, or are required, to provide reserve would be paid a standby fee, irrespective of whether a contingency actually occurs or not, and may receive little or nothing extra during a contingency. The overall expected energy price would equal the expected marginal cost of matching an increase in demand over all states.

These approaches would be expected to be equivalent, but the value which both dispatcher and dispatched are likely to place on ensuring some degree of certainty in this situation make the latter a more practical approach.

9.5.   An Optimal Reserve Targeting Pricing Model

We here present a simple illustrative example of an ORT pricing model. This model is similar to the model of Caramanis et al. (1987), though we consider a more specific form of reserve/generation trade-off function and instead of modelling a single energy conservation constraint with reserve requirements we use a different, but equivalent, formulation involving an energy conservation constraint for each state. The main original contribution of this and the following sections is in extending the analysis of the dual formulation.

---

[140]   Further constraints may be required but our aim here is simply to suggest a basic structure for further investigation.

We assume that the reliability of each generating unit is known and that the contingency states are independent apart from their links to the base state via the reserve requirements. For ease of discussion we ignore "electrical" constraints and variables. Hence our model revolves around the active power energy conservation constraint, in each contingency state, and the limits on generation and reserve availability. The requirement to cover lost generation capacity in a contingency state can be viewed as a simple analog of the requirement to respond to a frequency drop. While modelling the relationship between generation and reserve for each generating unit, we ignore the limits imposed on the reserve response of individual generating units by the wider system. For instance, some of the constraints we ignore include:

- Constraints imposed on a frequency keeping station. Such a station has the role of countering small frequency variations in normal operation. While the marginal generating unit can most economically fill this role, power systems are often operated with less expensive generating units taking this role, and so incurring the opportunity cost of not fully using their generation capacity. As discussed by Read and Ring (1995b), these stations can normally be modelled in a dispatch based pricing framework by fixing their output, producing a shadow price which accounts for the difference between that generating unit's marginal cost and the system marginal cost.

  The problem in the stochastic reserve situation is that the frequency keeping generating unit will be the first generating unit to respond to a frequency drop, but a model which minimises the cost of meeting contingencies will want to use cheaper reserve sources first. This would not be a problem if the smallest contingency required a response greater than the feasible response of the frequency keeping station, as we could then constrain it to always make the maximum response. The difficulty occurs when the frequency keeping station is capable of providing all the reserve capacity needed to cover a small contingency, and would do so in practice. We might avoid this problem, though, by defining a "contingency" to be any event large enough to require more than a frequency keeping response. Alternatively, we might modify the marginal reserve cost of the frequency keeping station so that it appears to be the cheapest source of reserve[141].

---

[141]   While this would violate the bounds set by the offer price of a normal market participant, a frequency keeping station is likely to be providing this service under contract to the dispatcher rather than responding to market signals.

- Some generating units, or curtailable loads, may have special links with particular pieces of equipment which may break down. This situation is analogous to that for the frequency keeping station, and might be handled in a similar way.

- Groups of generating units and curtailable load may be required to respond in a proportional manner so as to get a geographically dispersed response, reducing the need for dramatic changes in transmission line flows. This could easily be modelled in practice, simply by introducing a primal constraint and corresponding shadow price which reflect this linkage.

The model we present is consistent with that in Chapter 4 in that all the new terms introduced here can be represented by either the generalised variables or constraints of that model. These same facilities can also allow the modelling of many of the more complex effects which we ignore.

9.5.1. Notation and Terminology

We make use of the following sets:

- The Base State: A superscripted "0" indicates terms relating solely to the non-contingency state or "base state".

- The Set of Contingency States: The contingency states are denoted by the set *U*. This set does not include the base state.

- The Set of Generating Units: The set *PG* is used to denote the set of generating units. Demand is not included in this set as we assume that there is only one load.

The other terms we use are:

- π^0 is the probability that the system is in the base state, i.e. the state in which no contingencies have occurred.

- π^u is the probability of the system being in contingency state $u \in U$. These contingency states are only linked via the base state. Note that:

$$\sum_{u \in U} \pi^u = 1 - \pi^0 \qquad (9.1)$$

- P_D represents the single load, while generation is indicated by P_{Gi} $\forall i \in PG$. These variables are assumed to apply over all states, with additional variables or parameters being used to represent changes during contingency states.

- δ_i^u is a 0-1 parameter (note that it is <u>not</u> a variable) which takes a value of 1 if generating unit i is operational in state u, and a value of 0 if generating unit i fails in state u.

- p_{Ri}^u, p_C^u, and p_B^u. describe, respectively, the amount of spinning reserve (from generating unit i), load curtailment, and blackout actually utilised in contingency state u.

- P_{Ri} and P_C are respectively, the spinning reserve (from generating unit i) and load curtailment capacities which the dispatcher chooses to schedule in the base state. These place an upper limit on the amount of these resources which can be used during a contingency.

- P_{Ri}^{max} and P_C^{max} are, respectively the spinning reserve (from generating unit i) and load curtailment capacities made available to the dispatcher. These are exogenous variables and hence we treat them as constants. These values place an upper limit on the amount of these resources which the dispatcher can schedule in the base state.

- c_{Pi} is the fuel cost associated with generation and reserve response provided by generating unit i.

- c_C, and c_B are, respectively, the (constant) marginal costs of calling upon curtailable load and blackout. The blackout cost is assumed to be the system shortage cost, this exceeding all other costs we model.

- The "marginal generating unit" is that generating unit which reacts to satisfy a change in demand. Since demand is assumed to be the same in all states the marginal provider of energy is the same across all states.

- The "marginal reserve source", however, reacts to satisfy a change in contingency size in a given contingency state, and hence different reserve sources may fill this role in different contingency states. The concept of a marginal reserve source is redundant for the base state.

9.5.2. The Primal Formulation

A primal reserve/energy dispatch formulation is described by (9.2) to (9.11).

$$\underset{\substack{P_R^{PG} \; P_C \, , \, p_R^{PG \cup U} \, , \, p_C^U \, , \, p_B^U \geq 0 \\ P_G^{PG} \, , \, P_D}}{\textbf{Minimise}} \quad \sum_{i \in PG} \pi^0 c_{Pi} P_{Gi} + \sum_{u \in U} \pi^u \left(\sum_{i \in PG} \left(c_{Pi} \left(P_{Gi} + p_{Ri}^u \right) \delta_i^u \right) + c_C p_C^u + c_B p_B^u \right) \quad (9.2)$$

Conservation of Power

$$\left(\sum_{i \in PG} P_{Gi} \right) - P_D = 0 \qquad \qquad :\pi^0 \lambda_P^0 \qquad \qquad (9.3)$$

$$\sum_{i \in PG} \left(P_{Gi} + p_{Ri}^u \right) \delta_i^u + p_C^u + p_B^u - P_D = 0 \qquad :\pi^u \lambda_P^u \quad \forall u \in U \qquad (9.4)$$

Bounds on reserve usage

$$P_{Ri} \delta_i^u - p_{Ri}^u \delta_i^u \geq 0 \qquad \qquad :\pi^u \omega_{Ri}^u \quad \forall i \in PG \, \forall u \in U \quad (9.5)$$

$$P_C - p_C^u \geq 0 \qquad \qquad :\pi^u \omega_C^u \quad \forall u \in U \qquad (9.6)$$

Bounds on reserve availability and generation usage

$$-P_C \geq -P_C^{max} \qquad \qquad :\upsilon_C^+ \qquad \qquad (9.7)$$

$$-P_{Ri} \geq -P_{Ri}^{max} \qquad \qquad :\upsilon_{Ri}^+ \qquad \forall i \in PG \qquad (9.8)$$

$$-P_{Gi} - P_{Ri} \geq -P_{Gi}^{max} \qquad \qquad :\upsilon_{Pi}^+ \qquad \forall i \in PG \qquad (9.9)$$

$$P_{Gi} \geq P_{Gi}^{min} \qquad \qquad :\upsilon_{Pi}^- \qquad \forall i \in PG \qquad (9.10)$$

Load Settings

$$P_D = P_D^{set} \qquad \qquad :\beta_P \qquad \qquad (9.11)$$

The objective function (9.2) minimises the expected cost, over all states, of calling upon generation, spinning reserve, load curtailment and blackout. The δ_i^u parameter is used to eliminate the cost of generation and reserve of those generating units which have failed in contingency state u.

The base state energy conservation constraint is described by (9.3). During contingencies, (9.4) requires that the existing level of generation from the remaining operational generating units, plus any additional reserve they provide, equals demand less load curtailment and blackout. It is assumed that no operating generating unit will ever decrease its generation during a contingency, this being enforced by the requirement that $p_{Ri}^u \geq 0$. In effect, this means that the reserve carried in the base state, but unused in contingency state u must still be carried to cover any further contingencies which might occur but which have not been modelled.

Equations (9.5) and (9.6) specify the relationship between reserve usage and reserve scheduled for both spinning reserve and load curtailment. The δ_i^{su} parameter eliminates those relationships in (9.5) which are irrelevant due to the failure of generating unit i. Equation (9.4) plays an analogous role for blackout, as it limits blackout to being less than total demand. Equations (9.7) and (9.8) require that the capacities of spinning reserve and load curtailment scheduled in the base state cannot exceed the capacities made available by the market participants. Equation (9.9) limits the combination of generation and reserve capacity of generating unit i to be less than the maximum possible generation, while (9.10) places a lower bound on generation where appropriate. As in Chapter 4, the level of load is assumed to be externally specified, in this case by (9.11).

9.5.3. The Dual Formulation

The dual formulation is shown in (9.12) to (9.19).

$$\underset{\substack{\Box_P^{-PG},\Box_P^{+PG},\Box_R^{+PG}\geq 0 \\ \Box_R^{PG\cup U},v_C^+,\Box_C^U,v_B^+\geq 0 \\ \beta_P,\lambda_P^0,\Box_P^U}}{\textbf{Maximise}}\quad \beta_P P_D^{set} + \sum_{i\in PG}\left(v_{Pi}^- P_{Gi}^{min} - v_{Pi}^+ P_{Gi}^{max}\right) - v_C^+ P_C^{max} - \sum_{i\in PG} v_{Ri}^+ P_{Ri}^{max} \qquad (9.12)$$

Subject to:

Energy Price Defined by Demand Level

$$-\pi^0\lambda_P^0 - \sum_{u\in U}\pi^u\lambda_P^u + \beta_P = 0 \qquad\qquad :P_D \qquad\qquad (9.13)$$

Energy Price Constrained by Generation Costs

$$\pi^0\lambda_P^0 + \sum_{u\in U}\pi^u\lambda_P^u\delta_i^u - v_{Pi}^+ + v_{Pi}^- = \left(\pi^0 + \sum_{u\in U}\pi^u\delta_i^u\right)c_{Pi} \quad :P_{Gi} \quad \forall i\in PG \qquad (9.14)$$

Spinning Reserve Capacity Price Defined by rent on reserve usage

$$\sum_{u\in U}\pi^u\omega_{Ri}^u\delta_i^u - v_{Pi}^+ - v_{Ri}^+ \leq 0 \qquad\qquad :P_{Ri} \qquad \forall i\in PG \qquad (9.15)$$

$$\pi^u\lambda_P^u\delta_i^u - \pi^u\omega_{Ri}^u\delta_i^u \leq \pi^u c_{Pi}\delta_i^u \qquad\qquad :p_{Ri}^u \qquad \forall i\in PG, u\in U \quad (9.16)$$

Reserve Price/ Load Curtailment Capacity Charge
Ceiling Imposed by Spinning Reserve Cost

$$\sum_{u\in U}\pi^u\omega_C^u - v_C^+ \leq 0 \qquad\qquad :P_C \qquad\qquad (9.17)$$

$$\pi^u\lambda_P^u - \pi^u\omega_C^u \leq \pi^u c_C \qquad\qquad :p_C^u \qquad \forall u\in U \qquad (9.18)$$

Absolute Reserve Price Ceiling Imposed by Blackout Cost

$$\pi^u\lambda_P^u \leq \pi^u c_B \qquad\qquad :p_B^u \qquad \forall u\in U \qquad (9.19)$$

From (9.13) we can define the average energy price as:

$$\beta_P = \pi^0\lambda_P^0 + \sum_{u\in U}\pi^u\lambda_P^u \qquad\qquad (9.20)$$

Substituting this into (9.14) and making use of (9.1) yields:

$$\beta_P = c_{Pi} - \sum_{u\in U}\pi^u\left(\lambda_P^u - c_{Pi}\right)\left(1-\delta_i^u\right) + v_{Pi}^+ - v_{Pi}^- \quad \forall i\in PG \quad (9.21)$$

This states that, for generating unit *i*, the expected energy price equals its marginal cost less the profit foregone when it breaks down plus the rents on its generation capacity. If this unit is the marginal provider of energy, and never breaks down, the rents on its

generation capacity are zero[142] and the expected price equals its cost. Note that this does not mean this generating unit's fuel cost will equal the base state price, λ_P^0, but the expected price over all states. Thus it is possible for the marginal generating unit to make a loss (or profit) on P_{Gi} in the base state, its higher (lower) earnings on P_{Gi} during contingency states allowing it to just break even on average[143]. Typically, operating generating units will earn positive rents on their generation and reserve, and, as in Chapter 6, there is no particular reason why β_P should correspond to the marginal cost of any generating unit.

If $P_{Ri} > 0$ then, by complementary slackness, (9.15) becomes:

$$v_{Ri}^+ = \sum_{u \in U} \pi^u \omega_{Ri}^u \delta_i^u - v_{Pi}^+ \qquad\qquad \forall i \in PG \quad (9.22)$$

This equation defines the value of another unit of capacity taken to provide reserve in terms of the extra rent that can be earned from the additional reserve provision which this unit makes possible minus the loss due to withdrawing that unit of generation capacity.

If $p_{Ri}^u < P_{Ri}$ then the complementary slackness conditions applying to (9.5) require that $\omega_{Ri}^u = 0$, and (9.16) becomes $\lambda_P^u \leq c_{Pi}$. Further, if we also have $p_{Ri}^u > 0$ then complementary slackness requires that (9.16) must be binding, so $\lambda_P^u = c_{Pi}$. By a similar argument, if $p_{Ri}^u = P_{Ri}$ then complementary slackness requires that $\omega_{Ri}^u \geq 0$ and, since $p_{Ri}^u > 0$, (9.16) must be binding, i.e.[144]:

$$\pi^u \omega_{Ri}^u \delta_i^u = \pi^u \left(\lambda_P^u - c_{Pi} \right) \delta_i^u \qquad\qquad (9.23)$$

Since $\omega_{Ri}^u \geq 0$ (9.23) requires $\lambda_P^u \geq c_{Pi}$ for $\delta_i^u = 1$. Thus generating unit i will only provide reserve to contingency u if the marginal value of energy, λ_P^u, in that state equals or exceeds the marginal cost of generating unit i. If λ_P^u equals its marginal cost then generating unit i is the marginal reserve provider in that state, and makes no profit on its reserve capacity, i.e. $\omega_{Ri}^u = 0$. A profit is only made on reserve when generating unit i is not the marginal reserve provider and its reserve capacity is fully used, i.e. when $\lambda_P^u > c_{Pi}$ and $\delta_i^u = 1$. The total profit, over all states, which is earned on either generation capacity or reserve capacity, is given by substituting (9.23) into (9.22) to give:

---

[142] This does not imply that the rent on its reserve capacity will be zero, though. This rent is implicit in the definition of β_P in (9.20).

[143] Refer to the later discussion on the price distribution in contingency states.

[144] While the δ_i^u terms can be cancelled if we consider only the case where $\delta_i^u = 1$ we retain these terms so as to avoid confusion in later substitutions.

$$v_{Ri}^+ = \sum_{u \in U: \lambda_P^u \geq c_{Pi}} \pi^u \left(\lambda_P^u - c_{Pi} \right) \delta_i^u - v_{Pi}^+ \qquad\qquad \forall i \in PG \quad (9.24)$$

Similarly, if $P_C > 0$ then (9.17) gives the rent on scheduled curtailable load capacity as:

$$v_C^+ = \sum_{u \in U} \pi^u \omega_C^u \qquad\qquad (9.25)$$

Again, if $p_C^u < P_C$ then (9.6) requires that $\omega_C^u = 0$, and (9.18) becomes $\lambda_P^u \leq c_C$. If $0 < p_C^u < P_C$ then (9.18) becomes an equality with $\omega_C^u = 0$ and hence yielding $\lambda_P^u = c_C$. If, however, $p_C^u = P_C$ then $\omega_C^u \geq 0$ and (9.18) becomes:

$$\pi^u \omega_C^u = \pi^u \left(\lambda_P^u - c_C \right) \qquad\qquad (9.26)$$

This means that $\lambda_P^u \geq c_C$. Thus, like spinning reserve, load curtailment will be marginal, with $\omega_C^u = 0$, when $\lambda_P^u = c_C$. Load curtailment will be fully used, and its providers will make a profit of $\omega_C^u = \lambda_P^u - c_C$, if λ_P^u exceeds its marginal cost in state u. No load curtailment will be used when $\lambda_P^u < c_C$. The total profit, over all states, which is earned on load curtailment capacity is given by substituting (9.26) into (9.25) yielding:

$$v_C^+ \quad = \sum_{u \in U: \lambda_P^u \geq c_C} \pi^u \left(\lambda_P^u - c_C \right) \qquad\qquad (9.27)$$

Equation (9.19) requires that $\lambda_P^u \leq c_B$. Since the marginal cost of blackout is the system shortage cost we would expect blackout to be marginal when it is called upon. This can be seen by observing that if $p_B^u > 0$ then the complementary slackness conditions on (9.19) require $\lambda_P^u = c_B$.

This analysis demonstrates that in each contingency state the available reserve sources will be called upon in merit order, and that the marginal reserve provider sets the price of calling upon reserve in that state. The possibility of spinning reserve providers failing during contingency states means, however, that it is not necessarily true that the price of calling upon reserve in a given contingency state rises monotonically with the outage magnitude. For example, consider the situation where 1000MW of reserve is called upon in a contingency state in which no reserve sources fail. In this case generating unit i happens to be the marginal reserve provider. There might be another state in which generating unit i fails, but the total outage to cover is only 950MW. Now it might be necessary to call upon a more expensive reserve source than generating unit i, resulting in a higher price for reserve called upon despite the contingency being smaller[145].

---

[145]   If we were to define the "contingency" to be the sum of the loss of both generation and reserve capacity then we would expect the price of reserve called upon to rise monotonically with

9.5.4. Interpretation of the Base State Energy Price

An intuitive expectation of the behaviour of the λ_P^u terms might be that they are all greater than λ_P^0, on the grounds that less energy capacity is available in state u. This intuitive analysis overlooks the fact that the base state is constrained to have enough reserve to cover any contingency, while the contingency states may, depending on the formulation, have no reserve constraints imposed on them to cover further contingencies. Thus the base state is often the most tightly constrained. Consequently the behaviour of λ_P^0 may be quite different from that suggested by simple intuition. To illustrate this, we consider the ORT model for the case where a reserve providing generating unit is perfectly reliable and is operating above its lower generating bound. For this unit (9.21) requires:

$$\beta_P = c_{Pi} + \upsilon_{Pi}^+ \tag{9.28}$$

Substituting for υ_{Pi}^+ from (9.24) and for β_P from (9.20) we get:

$$\pi^0 \lambda_P^0 + \sum_{u \in U} \pi^u \lambda_P^u = c_{Pi} + \sum_{u \in U : \lambda_P^u \geq c_{Pi}} \pi^u \left(\lambda_P^u - c_{Pi} \right) - \upsilon_{Ri}^+ \tag{9.29}$$

This can be re-arranged as:

$$\pi^0 \lambda_P^0 = \pi^0 c_{Pi} + \sum_{u \in U : \lambda_P^u < c_{Pi}} \pi^u \left(c_{Pi} - \lambda_P^u \right) - \upsilon_{Ri}^+ \tag{9.30}$$

That is, an increase in the losses incurred by generating unit i in any of those states for which $\lambda_P^u < c_{Pi}$ causes λ_P^0 to rise in value, while an increase in the profits earned by generating unit i in any of those states for which $\lambda_P^u > c_{Pi}$ cause λ_P^0 to decrease in value. If we now consider the situation where this generating unit is the marginal generating unit, with $\beta_P = c_{Pi}$ (i.e., it is partially loaded with $\upsilon_{Pi}^+ = 0$), then (9.24) requires that:

$$\upsilon_{Ri}^+ = \sum_{u \in U : \lambda_P^u \geq c_{Pi}} \pi^u \left(\lambda_P^u - c_{Pi} \right) \tag{9.31}$$

Substituting this into (9.30), and replacing c_{Pi} with β_P yields:

$$\pi^0 \lambda_P^0 = \pi^0 \beta_P + \sum_{u \in U : \lambda_P^u < \beta_P} \pi^u \left(\beta_P - \lambda_P^u \right) - \sum_{u \in U : \lambda_P^u \geq \beta_P} \pi^u \left(\lambda_P^u - \beta_P \right) \tag{9.32}$$

---

contingency size. This observation is not useful in practice, though, because we need to know the contingency sizes, which are now a function of the available reserve capacity, before determining the optimal reserve capacity to have available.

This implies that if:

1. $$\sum_{u \in U: \lambda_P^u < \beta_P} \pi^u \left(\beta_P - \lambda_P^u \right) < \sum_{u \in U: \lambda_P^u \geq \beta_P} \pi^u \left(\lambda_P^u - \beta_P \right) \text{ then } \lambda_P^0 < \beta_P.$$

2. $$\sum_{u \in U: \lambda_P^u < \beta_P} \pi^u \left(\beta_P - \lambda_P^u \right) > \sum_{u \in U: \lambda_P^u \geq \beta_P} \pi^u \left(\lambda_P^u - \beta_P \right) \text{ then } \lambda_P^0 > \beta_P.$$

3. $$\sum_{u \in U: \lambda_P^u < \beta_P} \pi^u \left(\beta_P - \lambda_P^u \right) = \sum_{u \in U: \lambda_P^u \geq \beta_P} \pi^u \left(\lambda_P^u - \beta_P \right) \text{ then } \lambda_P^0 = \beta_P.$$

This result indicates that, given the prices for calling upon reserve in the contingency states, λ_P^0 can be viewed as a slack term which takes whatever value is required to ensure that adequate revenue is earned on base state generation. For instance, case 1 describes the situation where the expected marginal loss made by the marginal generating unit[146] in those contingency states for which $\lambda_P^u < \beta_P$ is less than the expected marginal profits earned in those contingency states for which $\lambda_P^u > \beta_P$, and hence λ_P^0 can be reduced below β_P while still allowing the owner of the marginal generating unit to break even. If the marginal generating unit is just breaking even then all other operating units must be earning a non-zero expected marginal profit. Conversely, in case 2, λ_P^0 must exceed β_P if the marginal generating unit is to break even when its expected marginal losses in contingency states exceed its expected marginal profits. In case 3 the losses match the profits so $\lambda_P^0 = \beta_P$.

For the smallest contingency we will generally have $\lambda_P^u < \beta_P$, but since it is possible for $\lambda_P^0 > \beta_P$ we can conclude that it is also possible for λ_P^0 to exceed some λ_P^u. That is, the price for reserve called upon during a contingency state can be lower than the base state price. This result contradicts the intuitive argument put forward above. Note that the price paid by firm load, P_D, is β_P, the expected price , because P_D appears in the base state and all contingency state equations.

9.6. A Graphical Representation of the ORT Pricing Formulation

Figure 9.1 depicts a simple representation of the outages which might occur. We assume that all spinning reserve sources are totally reliable, and do not fail in any contingency state[147]. The vertical axis describes the magnitude of possible outages, which

---

146   I.e., the marginal provider of energy across all states.

147   It is difficult to describe our results graphically when some reserve sources are only available for a subset of the contingency states.

we assume to be either a megawatt loss in power or a frequency drop, while the horizontal axis describes the cumulative probability of an outage equalling or exceeding a given magnitude. The outages represented may involve the simultaneous failure of several generating units. The base state is also shown, having no outage, though the probability of this state has been greatly under-represented in order to maintain a reasonable scale on the graph. Ring and Read (1994) refer to this representation as an "Outage Duration Curve" by analogy with the well known "Load Duration Curve", but note that it is the probability of each outage state, rather than the "duration" of any particular incident which is being represented.
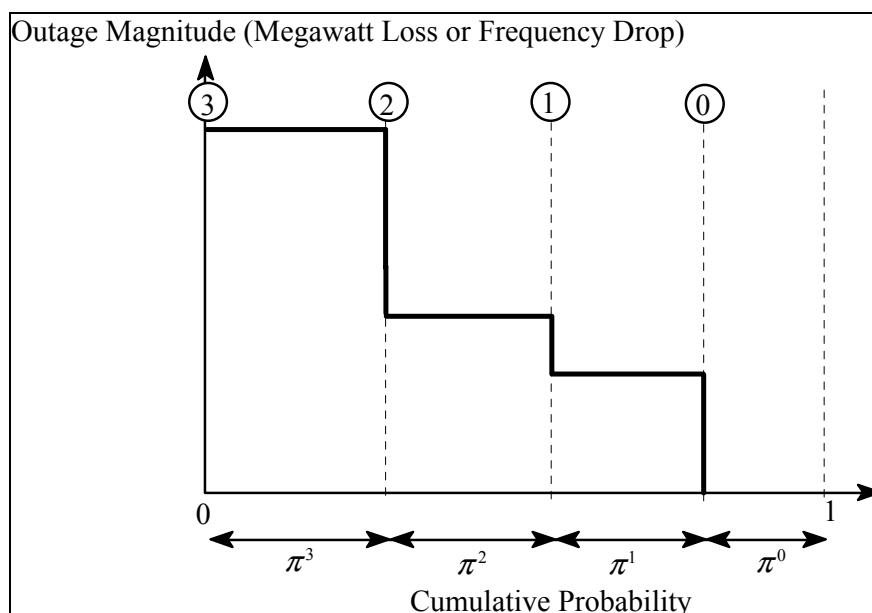
Figure 9.1: A graphical representation of the reserve requirements

The curve in Figure 9.1 can only be produced given the solution to the mathematical model above, as we do not know the output of the generating units, and hence the size of the outages, prior to solving that problem. Historic evidence could be used to produce an approximation, however, as is done by Anstine et al. (1963).

Figure 9.2 represents the pricing problem corresponding to Figure 9.1 for the case where there are two generating units providing spinning reserve and where we have the option of resorting to blackout. The cost of calling on each form of reserve in each contingency state is shown (the slope of each line is shown inside a circle), with the bold lines indicating the marginal reserve source in each contingency state. All reserve sources must provide reserve to cover the largest contingency, so the most expensive source, blackout, defines the marginal cost of reserve provision, λ_P^3, for that contingency. If this state were to occur with a probability of one, then the two generating units which provide reserve would make marginal profits for their reserve contribution equalling the difference

between λ_P^3 and their marginal costs. However, as the probability of being in this state is only π^3 these profits are weighted by this probability. Thus the point on vertical line 3 (i.e., the vertical axis) labelled $\pi^3 \lambda_P^3$ describes the expected value of a unit of reserve provided during a blackout. By subtracting the marginal costs of generating units 1 and 2 from this value, as indicated by the dotted sloping lines between vertical lines 2 and 3, we get the expected marginal profits during blackouts, $\pi^3 \omega_{R1}^3$ and $\pi^3 \omega_{R2}^3$ for generating units 1 and 2 respectively.

$$\omega_{R1}^3 = \lambda_P^3 - c_{P1}, \ \omega_{R2}^3 = \lambda_P^3 - c_{P2}, \ \omega_{R1}^2 = \lambda_P^2 - c_{P1}$$
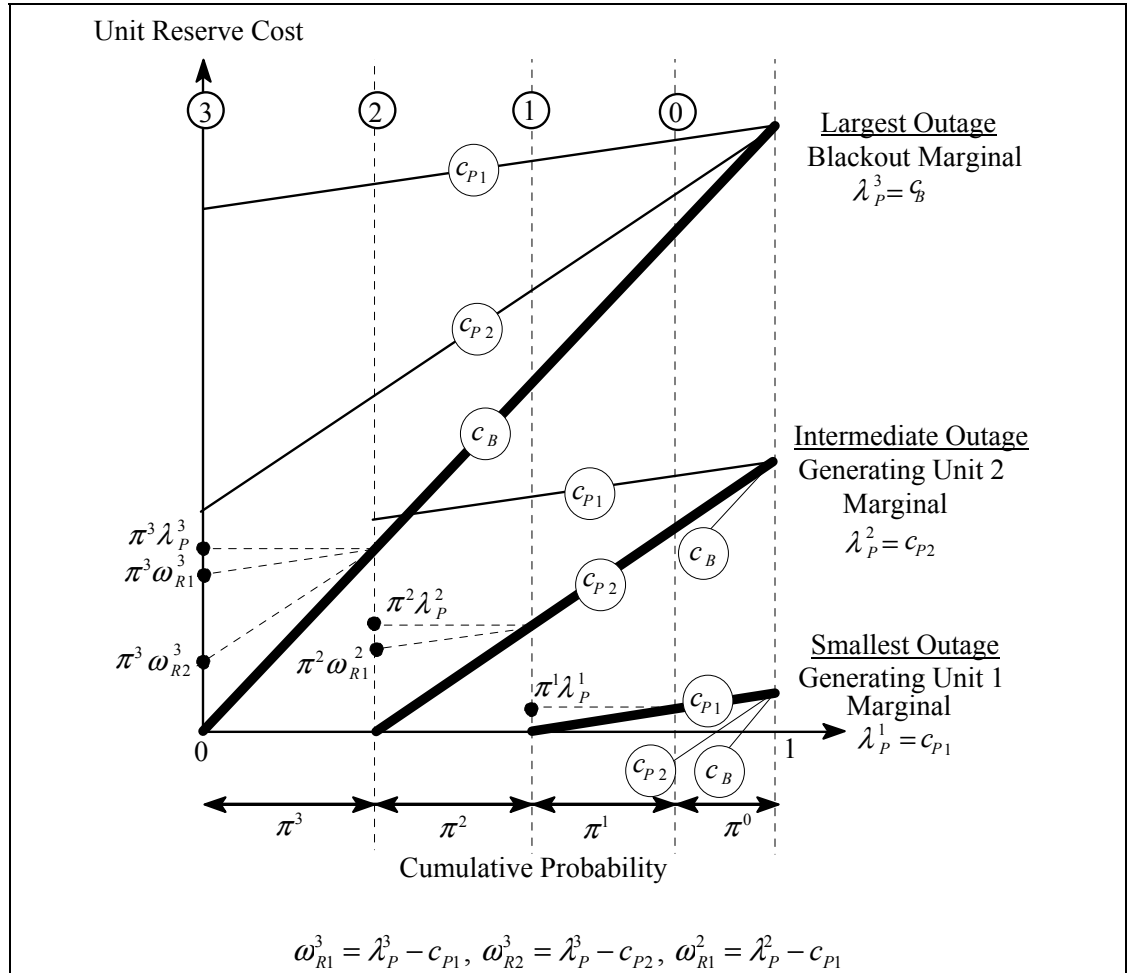
Figure 9.2: The marginal cost of calling upon reserve in each contingency state.

In state 2 blackout is not needed to cover the outage, so generating unit 2 is the marginal reserve source. The price for reserve called upon in this state is λ_P^2, and it occurs with probability π^2. Generating unit 1 is cheaper than unit 2, and so earns a profit of $\pi^2 \omega_{R1}^2$. Generating unit 1 is the marginal reserve source for the smallest contingency, and makes no profit in this state, while a loss would be made on the reserve contributions of unit 2 and blackout, were they to respond. Thus generating unit 1 makes a total marginal profit on reserve of $RP_1 = \pi^2 \omega_{R1}^2 + \pi^3 \omega_{R1}^3 = \pi^2 \left(\lambda_P^2 - c_{P1} \right) + \pi^3 \left(\lambda_P^3 - c_{P1} \right)$, while generating unit

2 makes a marginal profit of $RP_2 = \pi^3 \omega_{R2}^3 = \pi^3 \left(\lambda_P^3 - c_{P2} \right)$. No profit is made on blackout ($RP_B = 0$).
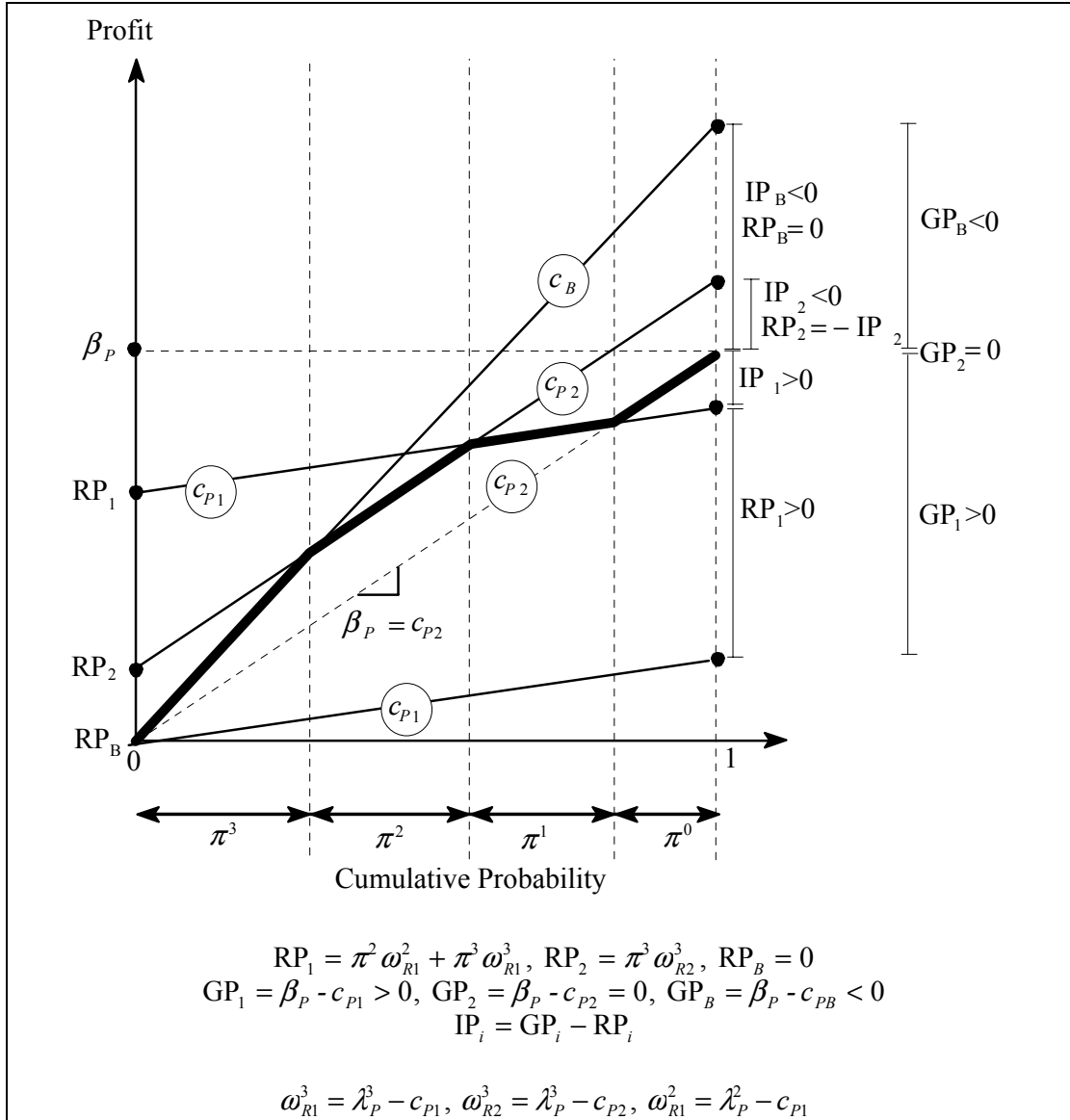
Figure 9.3: The profits on generation and reserve in each state.

Figure 9.3 depicts an alternative representation of this problem, but now provides information about the expected marginal profits earned on both energy and reserve capacity. The (solid) cost lines of each reserve source have been raised vertically so that the convex hull they form over the contingency states (indicated by the bold line for a cumulative probability less than $1 - \pi^0$) describes the cumulative, probability weighted, value of reserve provision for each point on the horizontal axis corresponding to a contingency state. The points where the cost lines intersect the vertical axis describe the expected marginal profit earned on reserve capacity. The bold line for state 0 has a slope

of λ_P^0, this slope taking what ever value is required to ensure that the expect prices over all states equals β_P (which, for this example, happens to equal c_{P2}), hence satisfying (9.20). To the right of the graph we again show the expected marginal profit on reserve, RP_i, but also show the expected marginal profit on base state generation over all states in which reserve is <u>not</u> called upon (including the base state), IP_i, and the sign of the total expected profit on base state generation, $GP_i = \beta_P - c_{Pi} = IP_i + RP_i$, for each reserve source.

How we use this representation depends on whether we are pricing by outcome or expectation. If pricing for outcome then we are only interested in the price of calling upon reserve in a given reserve state, or the price for energy in the base state, these values being defined by the slope of the bold line corresponding to that state. If, however, we are pricing based on an expectation then we are interested in the expected value of these prices.

The expected marginal reserve profit received by generating units 1 and 2 and blackout are $RP_1 = \pi^2 \omega_{R1}^2 + \pi^3 \omega_{R1}^3$, $RP_2 = \pi^3 \omega_{R2}^3$ and $RP_3 = 0$, respectively, these values corresponding to the point on the vertical axis where the cost lines of these reserve sources intersect.

Generating unit 2 is the marginal provider of energy because $\beta_{P2} = c_{P2}$. It is purely coincidental that $\beta_{P2} = \lambda_P^0$. Generating unit 2 makes no overall profit on its base state generation (i.e. $GP_2 = 0$) because while it earns a positive marginal reserve profit on this capacity in state 3 (i.e. $RP_2 > 0$), it incurs a matching loss in state 1 (i.e. $IP_2 = -RP_2 < 0$)[148]. Thus unit 2 only makes a profit on reserve provision and we would expect $P_{R2}^* = P_{R2}^{max}$, but being the marginal unit it must be able to respond to changes in demand, and hence we also require that $P_{G2}^{min} < P_{G2}^* < P_{G2}^{max} - P_{R2}^{max}$.

For generating unit 1 IP_1 describes the profit it earns on the energy it provides in the base state, this value being positive since $c_{P1} < \lambda_P^0$. Since this unit also makes a net profit in those states in which its reserve is called upon (i.e. $RP_1 > 0$), it must make a positive overall profit on its base state generation (i.e. $GP_1 > 0$). Thus generating unit 1 makes a profit on both its reserve capacity and its base state generating capacity. We would, therefore, expect this unit to be operated so as to maximise both its reserve capacity and base state energy output. This would correspond to the point $P_{R1}^* = P_{R1}^{max}$ and $P_{G1}^* = P_{G1}^{max} - P_{R1}^{max}$.

---

[148]   Generating unit 2 just breaks even in both states 2 and 0.

We have implicitly assumed that demand can always be meet in the base state by generation, hence there is no need for blackout to provide "base state generation". However, were blackout to be used in this way it would incur a loss in all states in which it was not called upon to provide reserve (states 0, 1, and 2), and hence $IP_B < 0$. Thus blackout provides no "generating capacity" as no profit can be made on such capacity, and it only breaks even on its reserve capacity (i.e. $RP_B = 0$).
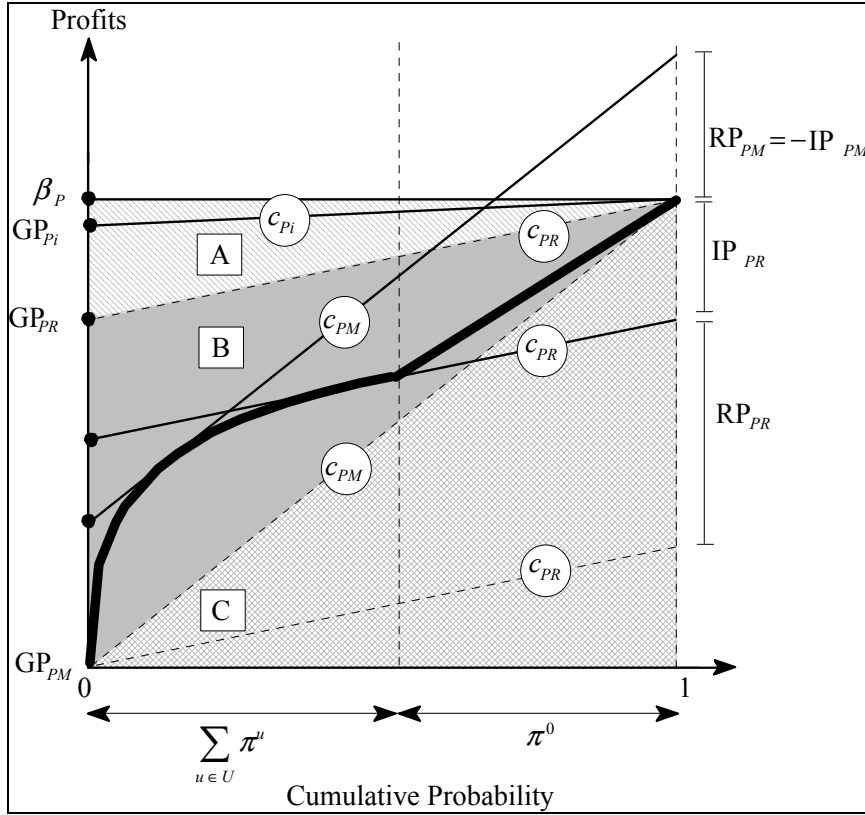
Figure 9.4: The trade-off between energy capacity and reserve capacity.

Figure 9.4 provides a summary of the factors which determine how the trade-off between reserve and energy provision is determined for a given power source. The slope of the bold line indicates the marginal value of reserve in each contingency state as well as the value of λ_P^0. Two solid lines are shown running at tangents to this bold line. The tangent with slope c_{PM} is the cost line for the marginal generating unit ($\beta_P = c_{PM}$). If we define $\lambda_P^{min} \equiv \min(\lambda_P^0, \lambda_P^u \ \forall u \in U)$ then the tangent with slope $c_{PR} = \lambda_P^{min}$ corresponds to the cost line for a (potentially hypothetical) generating unit which defines this price. For each of these generating units the expected marginal profit on base state generation (TR_{PR} and TR_{PM}) has been projected on to the vertical axis, effectively dividing the graph into three

regions. Introducing generating unit i, having an expected marginal profit on base state generation of TR_{PM} and a marginal cost of c_{Pi}, we can interpret these regions as follows[149]:

- Region A: If the line labelled c_{Pi} passes through this region then we must have $c_{Pi} < c_{PR}$. This means that generating unit i will make a profit in every states, and should therefore be fully loaded in all states. Thus this region corresponds to a unit which should only provide generating capacity ($P_{Gi}^* = P_{Gi}^{max}$, $P_{Ri}^* = 0$).

- Region B: If the line labelled c_{Pi} passes through this region then we must have $c_{PR} \leq c_{Pi} \leq \beta_P$. This unit can make a non-negative profit on both its reserve capacity and base state generation[150]. As no direct losses are incurred on reserve capacity, this capacity can be expected to be more profitable than generating capacity, and hence this region corresponds to a unit which should maximise its reserve capacity, and use its remaining capacity for generation ($P_{Ri}^* = P_{Ri}^{max}$, $P_{Gi}^* = P_{Gi}^{max} - P_{Ri}^{max}$[151]).

- Region C: If the line labelled c_{Pi} passes through this region then $c_{Pi} > \beta_P$. Unit i would not provide any base state generation capacity as it could not expect to make a profit by doing this. It can, however, make a profit on its reserve capacity in those contingency states in which $\lambda_P^u > c_{Pi}$, and will, therefore, maximise its reserve availability. Thus this region corresponds to a unit which only provides reserve generating capacity ($P_{Gi}^* = 0$, $P_{Ri}^* = P_{Ri}^{max}$)

---

[149] We ignore the possibility that generating unit i can break down.

[150] While it may make a loss on its base state generation in some states, this must be offset by profits earned in other states if we are to have $c_{Pi} \leq \beta_P$.

[151] Or $P_{Ri}^* = P_{Ri}^{max}$ and $P_{Gi}^* \leq P_{Gi}^{max} - P_{Ri}^{max}$ if $c_{Pi} = \beta_P$.
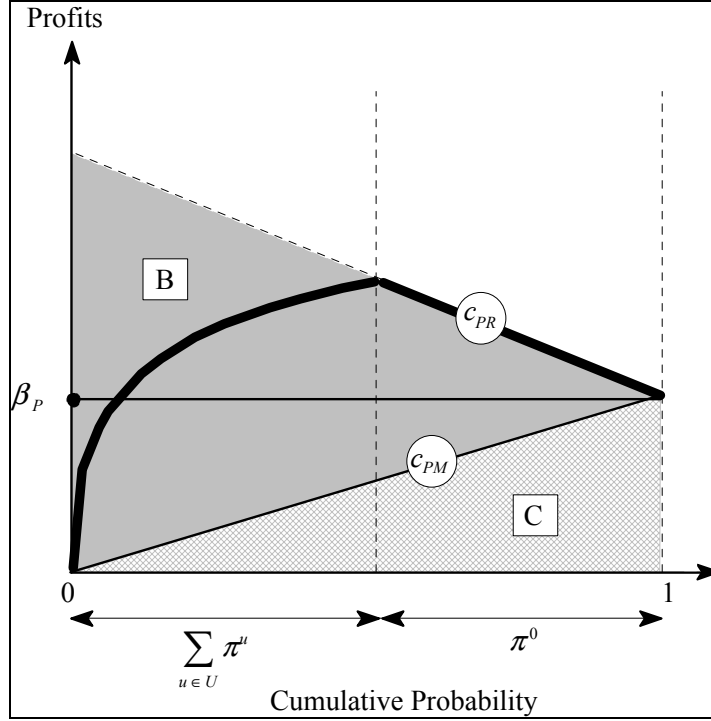
Figure 9.5: The special case when the base state price is negative.

Figure 9.5 illustrates the special case where $c_{PR} = \lambda_P^{min} = \lambda_P^0 < 0$. This corresponds to a situation where the expected profits on reserve capacity are so high that all energy sources (which we assume to have positive marginal costs) will be prepared to operate at a loss in the base case so as to be available to take advantage of the high contingency state prices. For this case region A will not exist, as no energy source can make a profit on base state generation.

9.7. Analysis of the Dual ORT Objective Function

Finally, an analysis of the dual objective function reveals the total expected profits. We can re-express the dual objective function in (9.12) in terms of observed primal solutions at the optimal dual prices. Observing that $P_D^* = P_D^{set}$ [152] and that:

- if $v_{Pi}^{-*} > 0$ then (9.10) requires $P_{Gi}^* = P_{Gi}^{min}$,

- if $v_{Pi}^{+*} > 0$ then (9.9) requires $P_{Gi}^* + P_{Ri}^* = P_{Gi}^{max}$,

- if $v_{Ri}^{+*} > 0$ then (9.8) requires $P_{Ri}^* = P_{Ri}^{max}$,

---

[152]   As usual, a "*" denotes an optimal solution value.

- if $\upsilon_C^{+*} > 0$ then (9.7) requires $P_C^* = P_C^{max}$.

The only non-zero terms in the function in (9.12) are:

$$\beta_P^* P_D^* + \sum_{i\in PG}\left(\upsilon_{Pi}^{-*} - \upsilon_{Pi}^{+*}\right)\left(P_{Gi}^* + P_{Ri}^*\right) - \upsilon_C^{+*} P_C^* - \sum_{i\in PG}\upsilon_{Ri}^{+*} P_{Ri}^* \qquad (9.33)$$

Or:

$$\beta_P^* P_D^* - \sum_{i\in PG}\left(\upsilon_{Pi}^{+*} - \upsilon_{Pi}^{-*}\right)P_{Gi}^* - \upsilon_C^{+*} P_C^* - \sum_{i\in PG}\left(\upsilon_{Pi}^{+*} + \upsilon_{Ri}^{+*}\right)P_{Ri}^* \qquad (9.34)$$

We observe from (9.21) that:

$$\upsilon_{Pi}^+ - \upsilon_{Pi}^- = \left(\beta_P - c_{Pi}\right) + \sum_{u\in U}\pi^u\left(\lambda_P^u - c_{Pi}\right)\left(1 - \delta_i^u\right) \qquad \forall i\in PG \qquad (9.35)$$

Substituting for $\upsilon_{Pi}^+ - \upsilon_{Pi}^-$ (from (9.35)), $\upsilon_{Pi}^+ + \upsilon_{Ri}^+$ (from (9.24)), and υ_C^+ (from (9.27)) in (9.33) yields:

$$\begin{aligned}
&\beta_P^* P_D^* - \sum_{i\in PG}\left(\beta_P^* - c_{Pi}\right)P_{Gi}^* + \sum_{i\in PG}\sum_{u\in U}\pi^u\left(\lambda_P^{u*} - c_{Pi}\right)\left(1 - \delta_i^u\right)P_{Gi}^* \\
&- \sum_{u\in U:\lambda_P^{u*}\geq c_C}\pi^u\left(\lambda_P^{u*} - c_C\right)P_C^* - \sum_{i\in PG}\sum_{u\in U:\lambda_P^{u*}\geq c_{Pi}}\pi^u\left(\lambda_P^{u*} - c_{Pi}\right)\delta_i^u P_{Ri}^*
\end{aligned} \qquad (9.36)$$

Or:

$$\begin{aligned}
&\beta_P^* P_D^* - \sum_{u\in U:\lambda_P^{u*}\geq c_C}\pi^u\left(\lambda_P^{u*} - c_C\right)P_C^* \\
&- \sum_{i\in PG}\left(\beta_P^* - c_{Pi}\right)P_{Gi}^* + \sum_{i\in PG}\sum_{u\in U}\pi^u\left(\lambda_P^{u*} - c_{Pi}\right)\left(1 - \delta_i^u\right)P_{Gi}^* \\
&- \sum_{i\in PG}\sum_{u\in U:\lambda_P^{u}\geq c_{Pi}}\pi^u\left(\lambda_P^{u*} - c_{Pi}\right)P_{Ri}^* + \sum_{i\in PG}\sum_{u\in U:\lambda_P^{u*}\geq c_{Pi}}\pi^u\left(\lambda_P^{u*} - c_{Pi}\right)\left(1 - \delta_i^u\right)P_{Ri}^*
\end{aligned} \qquad (9.37)$$

Here:

- $\beta_P^* P_D^*$ describes the total payment made by consumers for firm power,

- while $\displaystyle\sum_{u\in U:\lambda_P^{u*}\geq c_C}\pi^u\left(\lambda_P^{u*} - c_C\right)P_C^*$ is the effective discount received by those consumers who offer curtailable load.

- $\displaystyle\sum_{i\in PG}\left(\beta_P^* - c_{Pi}\right)P_{Gi}^*$ describes the expected profit on reliable base generation,

- while $\displaystyle\sum_{i\in PG}\sum_{u\in U}\pi^u\left(\lambda_P^{u*} - c_{Pi}\right)\left(1 - \delta_i^u\right)P_{Gi}^*$ is the expected profit foregone on base state generation when generating unit i fails. This might be viewed as the amount which needs to be paid by the owner of unit i to buy in reserve to cover its own generation

commitments where these commitments are based on this generating unit being perfectly reliable.

- $\sum_{i \in PG} \sum_{u \in U: \lambda_P^u \geq c_{Pi}} \pi^u \left(\lambda_P^{u*} - c_{Pi} \right) P_{Ri}^*$ is the expected profit on reliable reserve capacity,

- while $\sum_{i \in PG} \sum_{u \in U: \lambda_P^{u*} \geq c_{Pi}} \pi^u \left(\lambda_P^{u*} - c_{Pi} \right) \left(1 - \delta_i^u \right) P_{Ri}^*$ is the expected profit foregone on reserve

  capacity when generating unit *i* fails. As for base generation capacity, this might be viewed as the amount which needs to be paid by the owner of unit *i* to buy in reserve to cover its own reserve provision commitments where these commitments are based on this unit being perfectly reliable.

9.8. Pricing in the Longer Term

Our discussion thus far has only considered the immediate response to an outage. In the longer term the system must be re-dispatched to meet the changing situation. This introduces issues relating to inter-temporal constraints, integer variables, and sub-optimality. For instance, if we know in advance that a generating unit is going to fail then we can start ramping up a replacement generating unit well before the failure occurs. Without this perfect foresight the dispatcher might have to begin ramping up the new generating unit following the failure. During this ramping phase the system can be viewed as operating sub-optimally given that we now know we require this generating unit (see Figure 9.6).
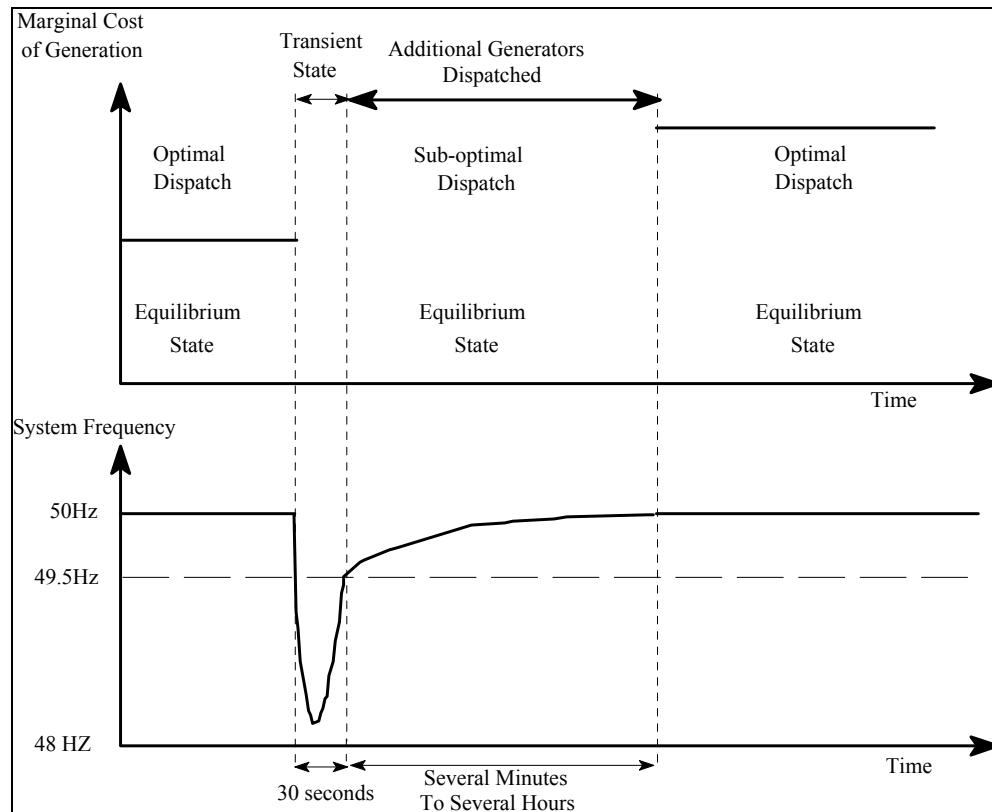
Figure 9.6: The behaviour of a power system during a contingency.

If we assume that a new pricing period begins when the system is re-dispatched, with reserve margins fully restored, following a contingency then these issues are outside the scope of the present discussion. Typically, the ramping generating unit, being hard against a constraint will not set the price. Instead the price will be set by some more flexible, but more expensive power source.

9.9. Conclusions

Theoretically, marginal cost based dispatch based spot prices can be derived for convex stochastic power system problems. In practice, though, the complexity involved in representing all facets of such problems means that some compromise in formulation detail is usually necessary. The key issue then, from the perspective of dispatch based pricing, is to determine a compromise formulation which corresponds adequately to the most important features of the real problem, and which also has a consistent economic interpretation. We suggest that in the context of the stochastic reserve problem the key goal in the primal problem should be to determine the optimal base state mix of reserve sources for the expected range of otherwise independent contingencies, and to require that, when a contingency occurs, any unused base state reserve capacity be maintained in case of further contingencies. This model effectively determines the optimal level of reserve to

be carried without blackout being called upon, hence we refer to it as an "Optimal Reserve Targeting", or ORT, model. While not included in our simple formulation, the requirement of feasible post-contingency transmission line flows would need to be included in a practical formulation.

We have analysed the pricing relationships corresponding to our ORT model formulation and have shown that dispatch based prices based on intentions should be consistent with prices based on outcomes. It is has been argued that when considering this model it is more intuitive to interpret the role of the dispatcher as optimising the system's automated response to contingencies rather than assuming that the dispatcher will determine the "optimal" response once a contingency occurs. We have also shown that the common assumption that the prices rise in contingency states relative to the base state is not necessarily true for this formulation[153]. The price in the base state takes what ever value is required to account for the difference between the marginal cost of meeting an increase in energy demand and the expected prices for reserve provision of all contingency states.

For the situation where generating units never breakdown we have shown that:

- if no state has a price exceeding a generating unit's marginal cost then that unit should use all its capacity to provide energy in the base state,

- if the generating unit's marginal cost equals or exceeds the price in at least one state, but is less than the marginal cost of meeting an increase in energy demand, then this unit should make the maximum possible reserve contribution while profitably using any remaining capacity to provide energy in the base state, while

- if the generating units marginal cost exceeds the marginal cost of meeting an increase in energy demand then this unit should make the maximum possible reserve contribution but make no base state energy contribution.

An analysis of the ORT pricing model objective function indicates that the actual market transactions can be partitioned into transactions in reliable power and reliable reserve capacity and transactions for power to cover each generating company's commitments where these commitments are based on the generating units being perfectly reliable.

---

[153]   Though this assumption should be true for some more general representation of the stochastic reserve problem.

The stochastic reserve model has strong inter-temporal aspects, as the system may take some time to fully adapt to the failure of lost generating unit. However, if a new pricing period begins once the systems initial response to a contingency has occurred then these issues are outside the scope of the present discussion.

Chapter 10

Extensions to Spot Pricing Theory:
Integer Dispatch Variables

10.1. Introduction

Integer variables are not consistent with a marginal cost pricing approach as the dispatcher only has the choice of whether or not to employ an integer process, incurring a cost which is only a function of that decision, not the amount produced. The commitment of a generating unit provides an example of an integer constraint; the unit is either on or off, and the act of committing it to run incurs a "start up cost[154]" while "running overhead" might be incurred for each period it is committed, both these costs being independent of the output level at which the unit is operated. Incurring these costs implies a commitment to participate in the market, and hence we will refer to these costs as *commitment costs*[155]. We distinguish between *short run commitment costs*, being costs for decisions which are made on similar time scale to the dispatch, and *long run commitment costs*, which relate to decisions based on much longer time scale considerations.

Here we concentrate on short run commitment costs which include:

- Generating unit start up costs. We define these costs so as to include the inevitable cost of shutting the generating unit down again.

- Generating unit "running overhead" costs. Both hydro and thermal generating units incur a fixed cost, either in terms of fuel usage or power consumption, for each period they are "on" irrespective of their output level.

- Curtailable load contracts. These may require that load be dropped in discrete bundles, implying a commitment cost in the exercising of the contract.

---

[154]   We define "start up" costs so as to include the inevitable shut down costs.

[155]   The use of the term "commitment cost" also occurs in the literature on strategic games. While we are not intending this terminology to necessarily be interpreted in the same sense, our definition is broadly analogous.

- Non-linear rents. These arise because not all costs associated with convex non-linear functions can necessarily be attributed to any single party.

Long term commitment costs, which are not considered here, may be associated with:

- Capital investment costs for generating companies and the transmission system.

- Generating station and transmission system maintenance costs.

- Miscellaneous overheads.

A significant implication of commitment costs is that there may not be a unique power price which is consistent with them. For instance, bringing a generating unit into the dispatch may sufficiently change the economics of the power system to make it appear to be uneconomical to dispatch that station. Consider the situation in which generating unit i is idle, having a fuel cost of 4c/kW, but could recover its commitment cost at the prevailing spot price of 5c/kW. Then it would appear that it should be dispatched. But suppose that, upon being dispatched, the generating unit which had previously defined the price is no longer required, generating unit i becomes marginal, and the marginal price falls to 4c/kW. Due to the commitment costs incurred this is now an unprofitable dispatch for generating unit i, although it may actually be the optimal solution. In fact there is no energy price which produces the optimal solution. Similar situations can arise with longer term investment problems. This "duality gap" problem, is not peculiar to dispatch based pricing, but is pervasive in integer optimisation problems, and can not be directly addressed within a standard marginal cost based pricing model. In most situations this does not matter, because the scale of the costs involved are small enough to cause a negligible distortion to the overall linearity (or more exactly convexity) of the problem, and there is enough uncertainty to obscure the effect. It is still desirable to have a pricing mechanism which accounts for duality gaps when they have a clear impact on the dispatch, though.

In the following section we review the literature on recovering integer costs. In the remainder of this chapter we then introduce some additional approaches for fixed cost recovery and consider how dispatch based pricing and best compromise pricing may be used to implement these new ideas as well as some of the ideas discussed in the literature review.

10.2. Review of Integer Pricing Models

Given the diverse range of commitment costs which can be encountered it may be necessary to consider more than one mechanism for recovering these costs. The best choice of mechanism to use in a particular instance will depend greatly on the magnitude of the commitment cost and which market participants incur that cost. For instance, Read and Sell (1989) use a DC approximation of the New Zealand system to estimate that only 10% to 30% of the costs of optimally sized transmission lines could be recovered through optimal short run pricing. They conclude that some form of fixed annual charge, in addition to normal spot prices, would be appropriate in this instance. Non-linear rents might also be covered by a fixed charge, as it is not possible to associate these with any single market participant. A positive rent will reduce the required magnitude of the fixed charge, while a negative rent will increase the required fixed charge. A fixed annual fee, used in conjunction with spot prices, can be interpreted as an example of non-linear pricing, an approach which allows pricing regimes to take their "optimal" form, rather than being required to satisfy some predetermined functional form[156]. That is, not only are the parameters of the pricing model optimised, but the functional form of the model is optimised. By forming the dual of a primal problem with the appropriate revenue constraints included we can determine the optimal pricing regime. The Priority Service Contract approach of Chao et al. (1986) and Chao and Wilson (1987), described in Chapter 2, is a form of non-linear pricing which aims to explicitly recover commitment costs (at the retail level) in an optimal manner[157]. Schweppe et al. (1988) report that an effective implementation of this approach may require a good knowledge of the benefit functions of market participants.

In contrast to transmission costs and non-linear rents, generating unit commitment costs can be associated with a single market player. If power systems were described by continuous variables, then, in the absence of economies of scale, standard spot prices would be able to recover the costs associated with socially optimal generation (and transmission) investment. For example, the results of Joskow (1976), Caramanis (1982), and the survey of Bergendahl (1983), suggest that if both capacity and actual production

---

[156]   Though for the example of transmission network long run commitment cost recovery this structure is still not necessarily truly optimal, but the fact that these long run costs are typically an order of magnitude greater than the cost of actually operating the network suggests that it is nearer to being an optimal regime than simply recovering these costs via the spot market.

[157]   The data requirements of this approach are very demanding though, with an accurate model of the dependence of each market participants behaviour on price being required if the true optimal pricing regime is to be found.

can be considered to be continuous variables, then the prices must obey the following equations[158]:

$$\text{If } P_{Gi}^t = 0 \text{ then } \beta_P^t < c_i + v_{Pi}^t, \text{ otherwise } \beta_P^t = c_i + v_{Pi}^t \quad \forall i \in I, t \in T \quad (10.1)$$

$$\sum_{t \in T} v_{Pi}^t = \alpha_i^t \kappa_i \qquad\qquad \forall i \in I \qquad (10.2)$$

$$\text{If } P_{Gi}^t = P_{Gi}^{max} \text{ then } v_{Pi}^t > 0, \text{ otherwise } v_{Pi}^t = 0 \qquad \forall i \in I, t \in T \quad (10.3)$$

Here v_{Pi}^t is the capacity charge for node i in period t and κ_i is the marginal cost of investment for generating unit i, while α_i^t represents the fraction of period t for which generating unit i is available. These equations indicate that, in the absence of other system constraints, generating unit i will only earn profits to cover its investment costs during those periods in which all its capacity is used.

In a market situation, companies will only commit to invest in generating capacity if they expect to recover their investment costs. This latter conclusion holds when the integer nature of investment is considered, but (10.2) must be replaced by a constraint, for each generating unit (or generating station), requiring that the net present value of the future stream of capacity charges equals or exceeds the investment cost. Equations (10.1) and (10.3) are included in standard spot pricing formulations, so one approach to this issue would be to use standard models to determine spot prices while leaving the generating companies to assess whether the expected capacity charges will recover their investment costs. It is unclear whether this approach will produce socially optimal investment.

An analogous situation applies to the recovery of short run commitment costs. If the spot prices based on marginal generation costs fail to recover commitment costs for a socially optimal dispatch, then we must consider modifying, or supplementing, the pricing regime. While the simplest way to recover commitment costs would be to add their average half hourly values to the spot price, such an unsophisticated approach would distort the spot prices, creating an asymmetry between generation and demand, and reducing the spot prices ability to provide favourable economic incentives to the market. The alternative of charging these cost components in the periods in which they occurred seems even worse.

Ramsey pricing (Baumol and Bradford, 1970) is an often suggested approach for modifying marginal cost based prices so as to satisfy revenue requirements while minimising the distortion to demand. Ansell and Lermit (1983) use Lagrangian techniques

---

[158]   For the situation of a single node, multi-period, multi-plant problem.

to produce what are effectively Ramsey prices for the New Zealand power pool for each half hour of a year[159]. While their model may not be practical in the context of real-time, or near real-time, spatially varying spot prices, it does provide useful information about the relative magnitudes of desirable energy and capacity charges, and could potentially be used to provide information about which types of period are well suited to price modifications. The need to know demand elasticities is often seen as an impediment to Ramsey pricing, and Schweppe et al. suggest that it might be simpler to minimise the sum of squares of price deviations. It should be noted, however, that an analogous result is achieved naturally by gaming generating companies, as they will aim to raise prices at those times when the market will tolerate them. Thus one possible solution is to simply let them exercise their market power to recover these costs where they can.

Vickrey (1979) suggests that if a generating unit were be committed to meet rising demand then its commitment could be delayed until the price rises to the point where the additional producer surplus produced by the price rise equals the commitment cost. This approach would require that the prices be updated very regularly, and would require a mechanism to return the extra producer surplus to the owner of the generating unit.

One of the best approaches reported in the literature for addressing commitment cost recovery has been proposed by S. Smith (1993). Smith proposes a heuristic methodology, based on an inter-temporal primal mixed-integer programming formulation, which determines prices, for a range of fixed daily scenarios, which ensure that the optimal mix of generating units will recover both their marginal and commitment costs over the time horizon. This model is intended for medium to long term analysis and consequently abstracts away from the dynamics of power system operation. A greedy algorithm is employed, based on the dual of the continuous relaxation of the primal, which works on the principle of raising price gradually in the period with the largest short-fall of supply. A generating unit will only be committed when prices have risen sufficiently that its commitment costs can be recovered during those periods in which the currently assigned prices exceed its marginal operating cost. A consequence of this method is that the first units to be introduced to the dispatch tend to be those with relatively low commitment costs though relatively high marginal costs. Because of their high marginal costs these stations would only be used around the peak. Sources with low marginal costs but high commitment costs are brought into the dispatch as the prices rise in periods surrounding the peak, and would tend to meet out the base load.

---

[159]   Strictly, they only estimate the spot price for every seventeenth half hour but this merely reflects the computation technology of the time.

Smith's model has the advantage of explicitly combining the dispatch process with the calculation of revenue recovering prices. Furthermore, it is computationally simple and fast, with Smith reporting solution times significantly faster than those for traditional Lagrangian relaxation methods. While Smith's method is an excellent planning aid, and provides good insights into the short run problem, it is limited in the context of dispatch batch based pricing by the limited detail involved in the modelling of network effects within a dispatch period and the system dynamics over several consecutive dispatch periods.

10.3. Spot Pricing Rent Manipulation

10.3.1. Adjusting the Pricing Model Objective Function

Another approach to commitment cost recovery would be to allow the dispatcher to collect extra revenue from the spot market. Read and Ring (1995b) argue that the dispatcher could do this by using a pricing model objective function which maximises the rents attributed to the transmission grid, hence minimising the need to recover commitment costs. This proposal has the advantage of not impacting on the interpretation of the nodal price differentials. It may not produce sufficient revenue to cover all commitment costs, though.

10.3.2. An Uplift Approach

For a single node problem we could apply an uplift by modifying the prices seen by either the generating companies or loads, so as to maintain consistency with their bids and offers, but not necessarily with the market clearing spot price. Figure 10.1. demonstrates such a situation. Prices for both generation and load can be moved independently within a range while not violating the bids or offers. The generation spot price is set to the lowest feasible level. Generating companies are paid at this price. Consumers pay this spot price to cover their usage of power, but face an additional unit charge, or uplift, which goes towards commitment cost recovery. This additional charge could be as great as the difference between the maximum and minimum possible spot prices.

Price

Nodal Supply Curve

Consumer Spot Price

Rent Goes Towards
Commitment Cost
Recovery

Generation Spot Price

Nodal Demand Curve

Quantity
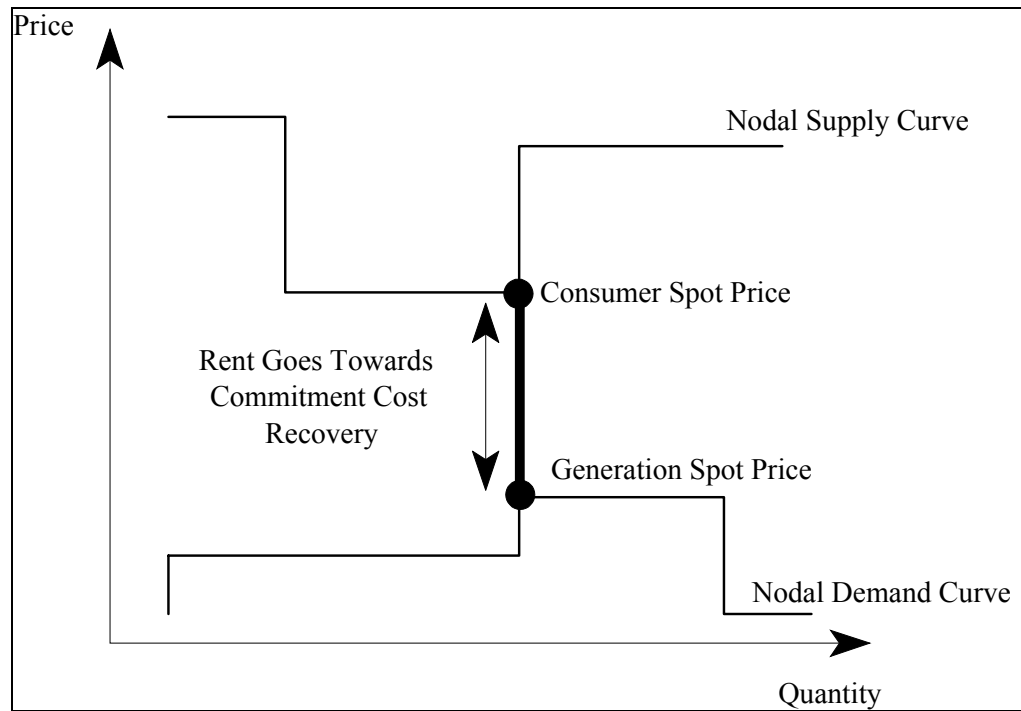
Figure 10.1: Recovering commitment costs when the spot price is defined in a range

    The situation depicted in Figure 10.1 will not be common in practice, but Figure 10.2 depicts a more typical situation. The consumer spot price can be raised above the generation spot price while maintaining consistency with the bids of consumers. The generation spot price is fixed by the offer and corresponds to the conventional market clearing spot price. The price difference produces revenue to cover commitment costs.
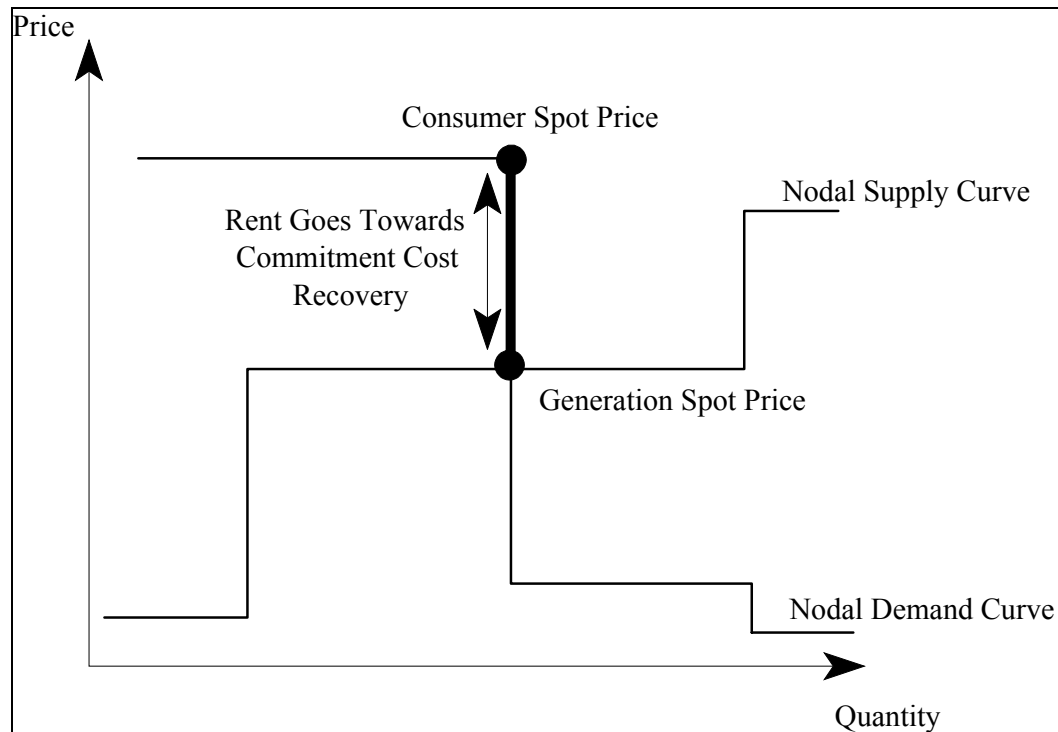
Figure 10.2: Recovering commitment costs when the spot price is unique

For a single node problem this approach is relatively straight-forward. However, the analogous multiple node problem is somewhat more difficult to address. For instance, if consumers and producers at each node see different prices then there is no longer a unique nodal price differential between each two nodes. One way around this would be to apply the uplift only at a reference node, with the normal market clearing spot price defining the generation spot prices but with the uplifted reference price, modified for inter-nodal losses and congestion, defining the consumption spot prices. This would ensure that all consumers (generators) were indifferent between their local price and the price of buying power at another node, paying the consumption (generation) spot price plus the cost of transmission. A drawback of this proposal is that the uplift at a node will be the uplift at the reference node modified by marginal losses and congestion rents, and this may unfairly attribute a greater or lesser proportion of the commitment costs to those consumers who are electrically distant from the reference node.

10.4. Application of Best Compromise Pricing

The method proposed by Smith provides a good, but simple, representation of both the primal unit commitment/dispatch problem as well as the economics of the corresponding pricing problem. The dispatcher may choose to use a model of this form to solve the primal problem, but our standard pricing formulation of Chapter 4 could not reproduce the prices of Smith's model, due to the integer nature of the problem and the lack

of any requirement that commitment costs be recovered. In this section we explore the ways in which best compromise pricing could be applied to this problem.

10.4.1. A Single Period Problem: Running Overhead Recovery
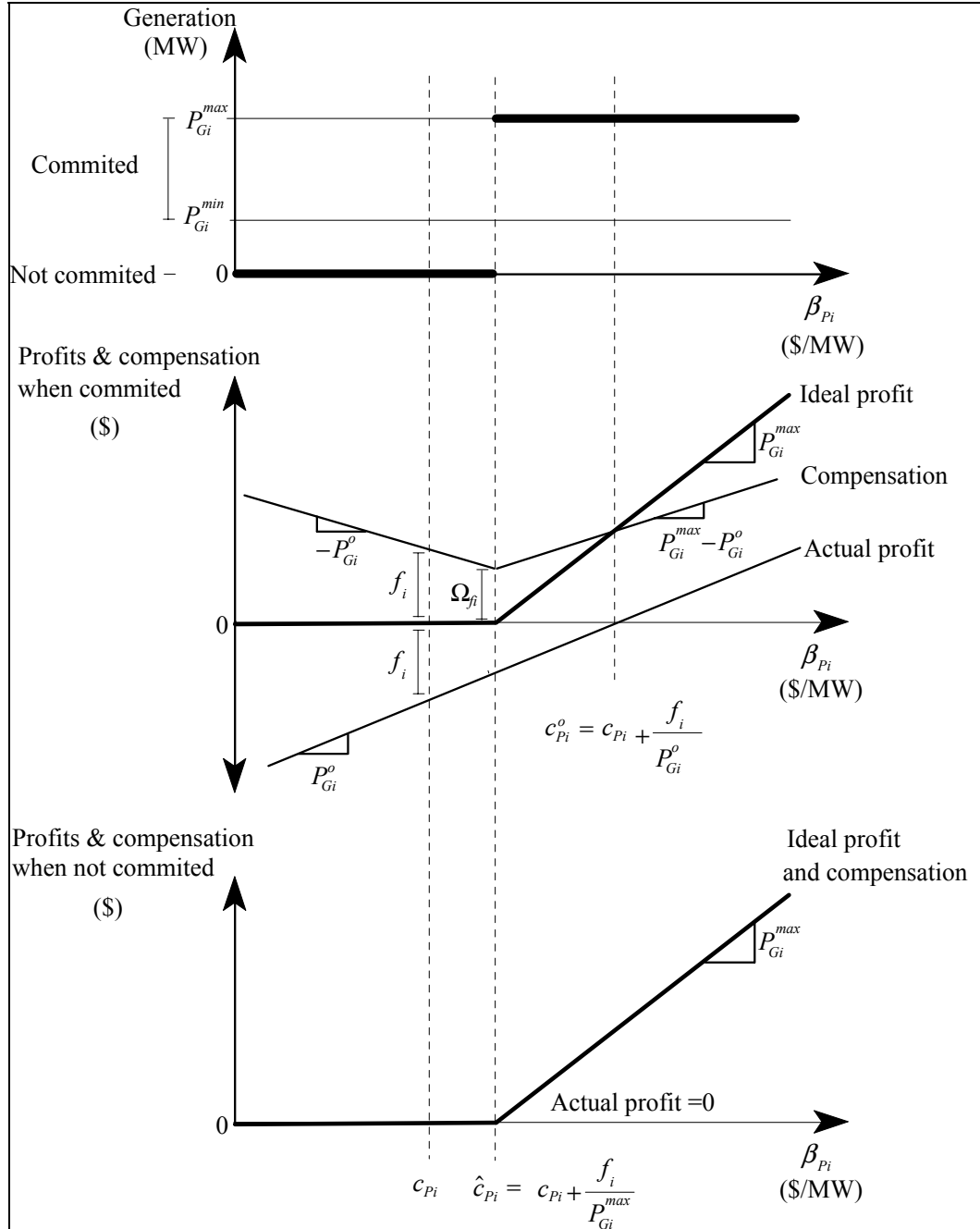
Figure 10.3: Best compromise pricing and commitment cost recovery

Figure 10.3 demonstrates how the best compromise pricing philosophy could be applied so as to account for commitment costs for a single period problem. Generating

unit i has a marginal cost of c_{Pi} and a running overhead of f_i[160]. The bold lines in the top graph show the optimal economic dispatch of unit i as a function of the energy price β_{Pi}[161] The unit should only be committed when β_{Pi} exceeds the minimum average cost, $\hat{c}_{Pi} = c_{Pi} + \dfrac{f_i}{P_{Gi}^{max}}$, because this is the lowest price which will cover all the costs incurred in committing and operating this unit. When committed, the owner of this unit will always wish that it be fully loaded[162]. This observation may not seem obvious, but it can be seen that the unit will only just break even if it generates P_{Gi}^{max} at a price equalling $\hat{c}_{Pi}$, and for any lower generation its profit will be negative, having a value of:

$$\left(\hat{c}_{Pi} - \left(c_{Pi} + \frac{f_i}{P_{Gi}^o} \right) \right) P_{Gi}^o = f_i \left(\frac{1}{P_{Gi}^{max}} - \frac{1}{P_{Gi}^o} \right) P_{Gi}^o$$
$$= \frac{f_i \left(P_{Gi}^o - P_{Gi}^{max} \right)}{P_{Gi}^{max}}$$
(10.1)

Alternatively, given an observed generation level of $P_{Gi}^o = 0.5 P_{Gi}^{max}$ then a consistent price would have to equal at least $c_{Pi} + \dfrac{2 f_i}{P_{Gi}^{max}}.$, which exceeds $\hat{c}_{Pi}$, and hence, if presented with the higher price, the owners of generating unit i would expect that unit to be fully loaded, and hence would expect to be compensated.

The second graph in Figure 10.3 illustrates the compensation required if generating unit i is committed (i.e. $P_{Gi}^{min} \leq P_{Gi}^o \leq P_{Gi}^{max}$). The ideal profit received by this unit is zero when β_{Pi} is less than $\hat{c}_{Pi}$ but rises by P_{Gi}^{max} for each unit increase in β_{Pi} beyond $\hat{c}_{Pi}$. If we define the actual average operating cost by:

$$c_{Pi}^o = c_{Pi} + \frac{f_i}{P_{Gi}^o}$$
(10.2)

Then the actual profit line has a slope of P_{Gi}^o and is negative while β_{Pi} is less than c_{Pi}^o (equalling $-f_i$ when $\beta_{Pi} = c_{Pi}$), but positive otherwise. The compensation required is just

---

160   This could equally be thought of as a start-up cost for a single period problem. We refer to it as a running overhead because we will develop a multi-period formulation, which also involves explicit start-up costs, below.

161   For a single period problem there is no reason to have the lower generating bound of a committed generating unit binding. If the price recovers all costs in the period it should run at full output, while if the price fails to recover all costs then it should not be committed. If the price exactly covers the costs then this unit is marginal, but, as discussed next, the marginal cost must be consistent with being fully loaded.

162   Given that the marginal operating cost is assumed to be constant across the entire generating range.

the ideal profit curve less the actual profit curve. The minimum compensation occurs when β_{Pi} equals $\hat{c}_{Pi}$, and has a value of:

$$\Omega_{fi} = \left(f_i + \left(\frac{f_i}{P_{Gi}^{max}} \right) \left(- P_{Gi}^o \right) \right) \delta_i$$

$$= \frac{f_i \left(P_{Gi}^{max} - P_{Gi}^o \right) \delta_i}{P_{Gi}^{max}}$$

(10.3)

Where:

$$\delta_i = \begin{cases} 1 & \text{if } P_{Gi}^o > 0 \\ 0 & \text{if } P_{Gi}^o = 0 \end{cases}$$

(10.4)

If the zero-one parameter δ_{fi} were to always equal one then if generating unit i had not been committed ($P_{Gi}^o = 0$) we would have $\Omega_{fi} = f_i$, even though unit i has not actually incurred the running overhead, hence δ_{fi} acts to set $\Omega_{fi} = 0$ in this event[163]. The value of Ω_{fi} can be thought of as compensation for under-utilisation of the committed capacity. Indeed, for a unit which is observed to have been committed, it has the same magnitude but the opposite sign to the loss defined in (10.1), and will therefore cancel out that loss. It also follows from (10.1) and (10.2) that[164]:

$$\Omega_{fi} = \left(\hat{c}_{Pi} - c_{Pi}^o \right) P_{Gi}^o$$

(10.5)

To interpret Ω_{fi}, consider the case where β_{Pi} equals $\hat{c}_{Pi}$. If the unit were fully loaded then $\hat{c}_{Pi}$ is the actual average operating cost and no compensation is required, and hence $\Omega_{fi} = 0$[165]. But if the unit is only partially loaded then the actual operating cost is c_{Pi}^o, which increases with decreasing P_{Gi}^o, and Ω_{fi} covers the difference between β_{Pi} and this cost. The value of Ω_{fi} also increases with decreasing P_{Gi}^o and its maximum possible value is:

$$\underset{P_{Gi}^o = P_{Gi}^{min} \to 0}{\text{Limit}} \quad \Omega_{fi} = f_i$$

(10.6)

---

[163]   The situation where the unit has not been committed is discussed below, and corresponds to the situation depicted in the lower diagram in Figure 10.3.

[164]   We do not require the term δ_{fi} here, because the situation where this should be zero corresponds to $P_{Gi}^o = 0$.

[165]   In this case the compensation shown in the middle graph of Figure 10.3 will be positive, with a slope of $-P_{Gi}^{max}$ for $\beta_{Pi} \leq \hat{c}_{Pi}$,, and be zero for $\beta_{Pi} \geq \hat{c}_{Pi}$.

The final graph in Figure 10.3 depicts the situation where generating unit i is not committed[166]. In this situation the actual profit is zero for all values of β_{Pi} and the optimal compensation equals the ideal profit.

It is apparent that regardless of whether or not generating unit i is committed, the compensation functions illustrated in Figure 10.3 have slopes of $-P_{Gi}^o$[167] for $\beta_{Pi} < c_{Pi}^o$ and $P_{Gi}^{max} - P_{Gi}^o$ for $\beta_{Pi} > c_{Pi}^o$. These slopes are identical to those of the best compromise pricing functions discussed in Chapter 7, provided we treat the generating unit lower bounds in the best compromise pricing model as being zero[168]. It follows that if all generating units with non-zero commitment costs are modelled as having lower generating bounds of zero and the "marginal" cost is replaced by the minimum average cost, $\hat{c}_{Pi}$, then the standard best comprise pricing formulation can determine spot prices which minimise the combined compensation required to recover the commitment costs and any costs due to mis-dispatch. The only difference is that the standard best comprise pricing formulation will set the compensation payment for generating unit i to be zero, rather than Ω_{fi} when $\beta_{Pi} = \hat{c}_{Pi}$. This can be remedied by adding Ω_{fi} to the compensation payments made to all committed generating units.

For a simple dispatch which only requires generation to equal a fixed demand level, (and hence consumers cannot be sub-optimally dispatched and demand therefore does not appear in the formulation)[169], the appropriate best compromise pricing formulation would be[170]:

---

[166]   Note that $P_{Gi}^o = 0$ in this case.

[167]   Though this is zero if the unit is not committed.

[168]   This is because if unit i should not be operating, but is, then it must be compensated for all its output, not just its output in excess of its lower generating bound. For a multi-period problem this will not necessarily be true, though.

[169]   If fixed demand was not assumed, with the prices required to be consistent with a demand curve of monotonically decreasing slope, then demand would appear in this formulation in a similar fashion to generation, except that the commitment costs for consumers will generally be zero.

[170]   As noted above, to ensure that a sub-optimally committed generating unit fully recovers its cost it is necessary to treat the lower generating bound to be zero regardless of whether or not the generating unit has been committed.

$$\sum_{i \in PG} \Omega_{fi} + \underset{\substack{v_P^{+PG} \, p_P^{-PG} \geq 0 \\ \beta_P}}{\textbf{Minimise}} \; \sum_{i \in PG} \left(v_{Pi}^{+} \left(P_{Gi}^{max} - P_{Gi}^{o} \right) + v_{Pi}^{-} \left(P_{Gi}^{o} - 0 \right) \right) \tag{10.7}$$

Subject to:

$$\beta_P - v_{Pi}^{+} + v_{Pi}^{-} = \hat{c}_i \qquad\qquad \forall i \in PG \tag{10.8}$$

$$\Omega_{fi} = \frac{f_i \left(P_{Gi}^{max} - P_{Gi}^{o} \right) \delta_{fi}}{P_{Gi}^{max}} \qquad\qquad \forall i \in PG \tag{10.9}$$

$$\delta_{fi} = \left\{ \begin{array}{ll} 1 & \text{if } P_{Gi}^{o} > 0 \\ 0 & \text{if } P_{Gi}^{o} = 0 \end{array} \right\} \qquad \forall i \in PG \tag{10.10}$$

We refer this formulation as a best compromise pricing with running overheads (RO) formulation. As it is a constant, the inclusion of Ω_{fi} in (10.7) does not impact on the optimisation, but means that the optimal objective function value describes the total compensation payment which needs to be made.

For this formulation Table 10.1 summarises the profits without compensation (PWC) that each generating unit will receive, the compensation payments that are required (CP), and the effective actual profits seen by each unit (EAP=PWC+CP). In each non-trivial case we have presented three equivalent expressions for each of PWC, CP, and EAP. In each instance the first expression is presented in terms of the non-zero terms which appear in (10.7), the second expression is in terms of the minimum average cost, $\hat{c}_{Pi}$, and the correction, Ω_{fi}, which accounts for the difference between this cost and the actual average cost, while the third term is in terms of c_{Pi} and f_i.

The first row in Table 10.1 shows that if the unit should be committed, and is committed, then the compensation accounts for the difference between the net profit actually earned at an output of P_{Gi}^{o} and the net profit which should have been earned had output been P_{Gi}^{max}. The results in the second row demonstrate that if the unit should be committed, but is not, then the compensation must equal the profits which would have occurred had it been committed with an output of P_{Gi}^{max}. The situation described in the third row is for the case where the unit should not be committed but is. The profit without compensation is negative (PWC<0), so the compensation payment has the same magnitude but opposite sign (CP=-PWC), producing the correct effective actual profit of zero (EAP=0). The final case corresponds to the situation where the unit should not, and was not, committed, and no compensation is required.

Commitment		Profit Without Compensation (PWC)	Compensation Payments (CP)	Effective Actual Profits (EAP=PWC+CP)
Optimal	Observed			
ON $v_{Pi}^+ \geq 0$ $v_{Pi}^- = 0$	ON $\delta_{fi} = 1$ $\Omega_{fi} \geq 0$	PWC $= v_{Pi}^+ P_{Gi}^o - \Omega_{fi}$ $= (\beta_P - \hat{c}_{Pi})P_{Gi}^o - \Omega_{fi}$ $= (\beta_P - c_{Pi})P_{Gi}^o - f_i$ ≥ 0	CP $= v_{Pi}^+ \left(P_{Gi}^{max} - P_{Gi}^o\right) + \Omega_{fi}$ $= (\beta_P - \hat{c}_{Pi})\left(P_{Gi}^{max} - P_{Gi}^o\right) + \Omega_{fi}$ $= (\beta_P - c_{Pi})\left(P_{Gi}^{max} - P_{Gi}^o\right)$ ≥ 0	EAP $= v_{Pi}^+ P_{Gi}^{max}$ $= (\beta_P - \hat{c}_{Pi})P_{Gi}^{max}$ $= (\beta_P - c_{Pi})P_{Gi}^{max} - f_i$ ≥ 0
ON $v_{Pi}^+ \geq 0$ $v_{Pi}^- = 0$	OFF $\delta_{fi} = 0$ $\Omega_{fi} = 0$	PWC=0	CR $= v_{Pi}^+ P_{Gi}^{max}$ $= (\beta_P - \hat{c}_{Pi})P_{Gi}^{max}$ $= \left(\beta_P - c_{Pi} - \dfrac{f_i}{P_{Gi}^{max}}\right)P_{Gi}^{max}$ ≥ 0	EAP $= v_{Pi}^+ P_{Gi}^{max}$ $= (\beta_P - \hat{c}_{Pi})P_{Gi}^{max}$ $= (\beta_P - c_{Pi})P_{Gi}^{max} - f_i$ ≥ 0
OFF $v_{Pi}^+ = 0$ $v_{Pi}^- \geq 0$	ON $\delta_{fi} = 1$ $\Omega_{fi} \geq 0$	PWC $= -v_{Pi}^- P_{Gi}^o - \Omega_{fi}$ $= (\beta_P - \hat{c}_{Pi})P_{Gi}^o - \Omega_{fi}$ $= (\beta_P - c_{Pi})P_{Gi}^o - f_i$ ≤ 0	CR=-PWC $= v_{Pi}^- P_{Gi}^o + \Omega_{fi}$ $= (\hat{c}_{Pi} - \beta_P)P_{Gi}^o + \Omega_{fi}$ $= \left(c_{Pi} + \dfrac{f_i}{P_{Gi}^{max}} - \beta_P\right)P_{Gi}^o$ $\qquad + \dfrac{f_i}{P_{Gi}^{max}}\left(P_{Gi}^{max} - P_{Gi}^o\right)$ ≥ 0	EAP=0
OFF $v_{Pi}^+ = 0$ $v_{Pi}^- \geq 0$	OFF $\delta_{fi} = 0$ $\Omega_{fi} = 0$	PWC=0	CR=0	EAP=0

Table 10.1: Summary of direct profit and compensation payments to unit *i* for a single period problem with running overheads.

It is insightful to consider the primal problem corresponding to this RO formulation. Defining the primal variable corresponding to (10.8) to be $-\Delta P_{Gi}$ and forming the "dual" of the continuous terms in (10.7) and (10.8) gives:

$$\textbf{Maximise}_{\square P_G^{PG}} \sum_{i \in PG} \left(-\hat{c}_{Pi} \Delta P_{Gi} \right) \tag{10.11}$$

Subject to:

$$\Delta P_{Gi} \leq P_{Gi}^{max} - P_{Gi}^{o} \quad : v_{Pi}^{+} \qquad \forall i \in PG \tag{10.12}$$

$$-\Delta P_{Gi} \leq P_{Gi}^{o} \qquad : v_{Pi}^{-} \qquad \forall i \in PG \tag{10.13}$$

$$\sum_{i \in PG} \Delta P_{Gi} = 0 \qquad\qquad :\beta_P \tag{10.14}$$

Following the best compromise pricing derivation of Chapter 7 we define $\Delta P_{Gi} = P_{Gi} - P_{Gi}^{o}$, and require that the observed generation equals demand, i.e. $\sum_{i \in PG} P_{Gi}^{o} = P_D$.

This allows us to re-write the primal as:

$$-\sum_{i \in PG} \left(c_{Pi} P_{Gi}^{o} + \frac{f_i}{P_{Gi}^{max}} P_{Gi}^{o} \right) + \textbf{Minimise}_{P_G^{PG}} \sum_{i \in PG} \left(c_{Pi} P_{Gi} + \frac{f_i}{P_{Gi}^{max}} P_{Gi} \right) \tag{10.15}$$

Subject to:

$$0 \leq P_{Gi} \leq P_{Gi}^{max} \qquad\qquad \forall i \in PG \tag{10.16}$$

$$\sum_{i \in PG} P_{Gi} = P_D \tag{10.17}$$

If we ignore the constants in (10.15) this is just the linear relaxation of the actual integer problem[171]. More generally, (10.15) equals the difference in costs between the optimal solution of the relaxed problem and the relaxed costs of the observed solution. Since (10.15) must equal the optimised terms in (10.7), we can conclude that the compensation provided by the RO formulation equals the compensation required for a linear relaxation of the integer problem, plus an extra component (the Ω_{fi} terms) to account for the difference between the actual average cost of the observed dispatch and the minimum average cost. The minimum compensation will occur when $P_G^{o} = P_G^{*}$, but this compensation and the corresponding best compromise market clearing prices will not be consistent with the optimal integer dispatch unless P_G^{*} happens to correspond to the optimal integer dispatch.

This discussion demonstrates that the RO formulation can readily account for single period commitment cost problems by producing a set of energy prices (β_P), per unit compensation payments (v_{Pi}^{+}, v_{Pi}^{-}), and lump sum compensation (Ω_{fi}) which ensure that all

---

171   Note that any generation lower bounds that might exist in the actual integer problem do not appear here.

parties are satisfied at minimum cost. If the dispatch is optimal, compensation will only be made where necessary to ensure that the owners of (potential) marginal generating units are (not) happy to enter the dispatch.

10.4.2. A Multiple Period Problem: Start up and Running Overhead Cost Recovery

The general multi-period commitment cost problem assumes that a start up cost is incurred in the period when a unit is committed, while running overheads are incurred in each subsequent period for which it remains committed. Some features which make this general problem difficult, from a best compromise pricing perspective, are:

- Unlike the example in Figure 10.3, whether a generating unit is committed or not in a given period does not depend solely on the profits that could be earned in that period, but on the profits that could be earned over all the adjacent periods in which the unit is committed. This means that there is strictly no unique "minimum average cost" with which to replace each generating unit's marginal cost in the best compromise pricing formulation.

- Furthermore, if best compromise pricing is to deem a generating unit to be committed sub-optimally, then the ideal compensation is the difference between what this unit would earn for its optimal commitment and what it earned for its observed solution. Determining what the optimal commitment should be is an integer problem.

Thus to fully address this problem we might need to resort to a multi-period, integer formulation similar to that proposed by Smith. There are, however, several ways in which the standard best compromise pricing model, or a simple multi-period extension of it, might be used to approximately address the general commitment cost problem. As suggested in the literature review, we could allow generating companies to include their commitment costs in their offers, and then use conventional best compromise pricing. This is in fact precisely what the single period model of Figure 10.3 implies they should do. In reality, though, this approach may result in generating units being committed and shut down more often than would occur if the dispatcher were responsible for ensuring that the start up costs were recovered. As a consequence, the owners of these units might have to increase their offer prices considerably to offset the high rate of incurring start up costs. The result may be inflated prices which, while consistent with the bids and offers, produce a lower welfare level than would arise from a socially optimal approach to this problem.

Alternatively, the dispatcher could ignore the commitment costs and use the standard best compromise pricing formulation to determine the prices for each individual period in

isolation, and then, given these prices[172], partition the planning horizon into "dispatch epochs" consisting of contiguous periods during which each generating unit should be either committed or not committed. While we ignore how this partition is determined, we suggest that a small dynamic programming formulation, possibly of the type used by Muckstadt and Koenig (1977), could fill this role[173]. The resulting commitment schedule could be compared with the actual commitment, which may include a different set of dispatch epochs, to identify apparent inconsistencies in the observed dispatch, and compensation could be derived based on these inconsistencies. The precise method and rules for determining this compensation are not obvious, and would need to be negotiated.

Yet another approach is to use a best compromise pricing model based on the RO formulation. While it is not clear how such a model could be applied to a general start up cost recovery problem, we here endeavour to show how it can be applied for a simplified example. So as to abstract away from the complexities of identifying all combinations of actual and ideal dispatch epochs, we consider a simple multiple period problem covering a single cycle in demand, which we assume to be a day. The simplifying assumptions we make are:

- The optimal prices rise monotonically to the daily peak and then fall monotonically.

- We assume that all units are off both before and after this day.

- Given the monotonic behaviour of the prices, the optimal commitment of each generating unit will involve, at most, incurring a single start up cost during the cycle. We assume that the observed commitment obeys this condition as well.

---

[172]   The compensation implied by the best compromise pricing solution can be ignored in the approach we propose here.

[173]   A dynamic program could be solved for each generating unit in isolation, with two states (committed or uncommitted), and a number of stages equalling the number of periods.

CASE 1

B+C+D≥ F_i so generating unit i should be committed.
Per unit generation compensation for Commitment 1 = B+D
Per unit generation compensation for Commitment 2 = 0
Per unit generation compensation for Commitment 3 = A+E

CASE 2

B+C+D< F_i so should not be committed
Per unit generation compensation for Commitment 1 = F_i-C
Per unit generation compensation for Commitment 2 = F_i-(B+C+D)
Per unit generation compensation for Commitment 3 = A+E+(F_i-B+C+D)

Figure 10.4: Compensation for a simplified multi-period problem

Case 1 in Figure 10.4 depicts the resulting problem for the situation where generating unit i should be committed (i.e. B+C+D≥ F_i). Generating unit i should ideally be committed for the time period labelled "Commitment 2" as its profits, represented by the area B+C+D, exceed its start up cost F_i. If "Commitment 1" occurs instead, then the owners of unit i earn C, but must be compensated by B+D, because, having been committed, the unit should have been run for these periods and earned these rents. If "Commitment 3" occurs, then compensation of A+E must be paid to cover the losses incurred by operating unnecessarily in unprofitable periods.

Finally, if the unit was not committed at all it should be compensated by B+C+D minus F_i, because the start up costs will not been incurred in this case. Note that compensation for the running overhead is implicitly accounted for by using $\hat{c}_{Pi}$, rather than

c_{Pi} or c_{Pi}^o as the reference cost. If the unit is not generating at full output when optimally committed, then compensation would be provided as for the single period problem discussed above.

Case 2 corresponds to the situation where B+C+D does not cover the start up cost of generating unit *i*. Now the optimal commitment is to leave the unit idle for the whole cycle, and no compensation is required in that case. If Commitment 1 occurs then the operating profit of C fails to cover the commitment cost. Were the owners of this generating unit to be given compensation of B+D then they still would not recover the start up cost, hence the correct action is to conclude that this unit was committed erroneously and pay compensation of F_i-C so as to exactly cancel out the losses incurred by the owners. For Commitment 3 compensation of A+E must be paid to cover the cost of being dispatched in periods in which the price is less than the effective cost, while an additional F_i-(B+C+D) must be paid to cover that part of the start up cost not covered by the profits within the period. For Commitment 2, there will still be a loss of F_i-(B+C+D), and this must be paid in compensation, even though that is the optimal dispatch, <u>given</u> that the unit was (wrongly) committed for the day.

An intuitively appealing best compromise pricing formulation for this problem would be similar to the RO formulation, but would include additional terms which cover the shortfall if commitment costs are not recovered for units which are operating, and which cancel out the normal compensation if the unit did not operate, and should not operate. The additional terms we require are developed in the following discussion.

Ignoring start up costs, the expected profits over all periods for an optimally dispatch generating unit should be:

$$\text{Expected profit without start up costs} = \sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max} \qquad (10.18)$$

With start up costs, though, the expect profit becomes:

$$\text{Expected profit with start up costs} = \begin{cases} 0 & \text{if } \sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max} - F_i < 0 \\ \sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max} - F_i & \text{if } \sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max} - F_i \geq 0 \end{cases} \qquad (10.19)$$

By inspection of Table 10.1 it can be surmised that a multiple period version of the RO formulation, which we refer to as a MRO formulation[174], will reproduce these expected

---

[174]   This formulation will minimise the sum of the individual period compensations subject to all the constraints which apply in each period. The compensation it provides will be identical to the cumulative compensation determined by solving the BCPRO formulation for each period.

profits provided that the unit is correctly committed (i.e., it is committed and should be, or it is not committed and should not be). Modifications would need to be made to such a formulation for it to determine the correct compensation in those situation where the commitment is not optimal. There are two cases which need to be considered:

1. If $\sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max} - F_i \geq 0$, but the unit is not committed, then the MRO formulation will give compensation of $\sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max}$. Thus the MRO formulation must be modified so as to subtract F_i from this to take account of the fact that the unit did not incur a commitment cost.

2. If $\sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max} - F_i < 0$, and the unit is committed, the MRO formulation will provide enough compensation so that the unit earns an effective profit of $\sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max}$, when it should in fact earn none. Thus extra compensation of $F_i - \sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max}$ needs to be included in the MRO formulation, F_i to cover the fixed cost incurred and $-\sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max}$ to cancel out the effective profit implied by the standard model.

We account for these extra compensation requirements by introducing the term $U_i - F_i (1 - \delta_{Fi})$ into the MRO objective function, where:

$$\text{maximum} \left\{ \begin{array}{c} 0 \\ F_i - \sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max} \end{array} \right\} \leq U_i \qquad (10.20)$$

$$\delta_{Fi} = \left\{ \begin{array}{ll} 1 & \text{if unit } i \text{ commited at some time} \\ 0 & \text{if unit } i \text{ never commited} \end{array} \right\} \qquad (10.21)$$

The resulting formulation is:

$$\sum_{i \in PG}\left(\sum_{t \in T}\left(\Omega^t_{fi}\right) - F_i\left(1 - \delta_{Fi}\right)\right) + \underset{\substack{\Box^{+PG \cup T}_P, \Box^{-PG \cup T}_P, U^{PG} \geq 0 \\ \Box^T_P}}{\textbf{Minimise}} \sum_{i \in PG}\left(\sum_{t \in T}\left(v^{+t}_{Pi}\left(P^{max}_{Gi} - P^{ot}_{Gi}\right) + v^{-t}_{Pi}\left(P^{ot}_{Gi} - 0\right)\right) + U_i\right)$$

$$(10.22)$$

Subject to:

$$\beta^t_P - v^{+t}_{Pi} + v^{-t}_{Pi} = \hat{c}_i \qquad\qquad \forall i \in PG, \forall t \in T \quad (10.23)$$

$$\sum_{t \in T} v^{+t}_{Pi} P^{max}_{Gi} + U_i \geq F_i \qquad\qquad \forall i \in PG \qquad\qquad (10.24$$

$$\Omega^t_{fi} = \frac{f_i\left(P^{max}_{Gi} - P^{ot}_{Gi}\right)\delta^t_{fi}}{P^{max}_{Gi}} \qquad \forall i \in PG, \forall t \in T \quad (10.25)$$

$$\delta_{fi} = \begin{cases} 1 & \text{if } P^{to}_{Gi} > 0 \\ 0 & \text{if } P^{to}_{Gi} = 0 \end{cases} \qquad \forall i \in PG, \forall t \in T \quad (10.26)$$

$$\delta_{Fi} = \begin{cases} 1 & \text{if unit } i \text{ commited at some time} \\ 0 & \text{if unit } i \text{ never commited} \end{cases}$$

$$\forall i \in PG \qquad\qquad (10.27)$$

We refer to this as a multiple period best compromise pricing with start up costs and running overheads (MSRO) formulation. For ease of exposition we refer to those terms in (10.22) other than $U_i - F_i\left(1 - \delta_{Fi}\right)$ as the MRO compensation terms. The MSRO formulation works to minimise the MRO compensation terms plus U_i (note that $F_i\left(1 - \delta_{Fi}\right)$ is constant). It is apparent that:

- If $F_i - \sum_{t \in T} v^{+t}_{Pi} P^{max}_{Gi} \leq 0$ then unit i should be committed, and the MSRO formulation will set $U_i = 0$. If the unit is not fully loaded in the periods in which it should be committed, or is loaded in periods when it should not be committed, then compensation will be provided by the MRO compensation terms. If $\delta_{Fi} = 1$ then $U_i - F_i\left(1 - \delta_{Fi}\right) = 0$ because the unit is optimally committed. If, however, $\delta_{Fi} = 0$ (this is case 1 discussed above) then $U_i - F_i\left(1 - \delta_{Fi}\right) = -F_i$, this term subtracting the commitment cost from the MRO compensation[175].

- If, however, $F_i - \sum_{t \in T} v^{+t}_{Pi} P^{max}_{Gi} > 0$, then unit i should not be committed, and the MSRO formulation will set $U_i = F_i - \sum_{t \in T} v^{+t}_{Pi} P^{max}_{Gi}$. If $\delta_{Fi} = 1$ (this is case 2 discussed above)

---

[175]   The MRO compensation will include compensation for the periods when the unit is operating above its lower bound when $v^{-t}_{Pi} \geq 0$ as well as compensation for the periods when the unit is not fully loaded when $v^{+t}_{Pi} \geq 0$.

then $U_i - F_i(1 - \delta_{Fi}) = F_i - \sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max}$, the F_i term covers the start up cost incurred

by the owner of unit i while the $-\sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max}$ cancels out the matching term

appearing in the MRO compensation (but does not cancel out the $\sum_{i \in PG} v_{Pi}^{-t} P_{Gi}^{ot}$ terms

which compensates the owner of unit i for the losses incurred in those periods where,

even if there were no start up cost, the unit should not have operated). If, however,

$\delta_{Fi} = 0$ then $U_i - F_i(1 - \delta_{Fi}) = -\sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max}$,   this   term   cancelling   out   the

compensation of $\sum_{t \in T} v_{Pi}^{+t} P_{Gi}^{max}$ determined by the MRO compensation terms.

It is apparent, therefore, that the MSRO formulation will determine the compensation
payment required so as to ensure that (10.19) is satisfied.

We now consider the primal form of the MSRO model. We define $-\Delta P_{Gi}^t$ to be the
shadow price on (10.23) and $1 - z_i$ to be the shadow price on (10.24). We also put the
constant terms relating to start up costs $(-F_i(1 - \delta_{Fi}))$ in (10.22) into the primal objective
function[176]. The corresponding primal formulation is:

$$\sum_{i \in PG} -F_i(1 - \delta_{Fi}) + \mathbf{Maximise}_{1-\mathbf{z}^{PG} \geq 0,\square P_G^{T \cup PG}} \sum_{i \in PG} \left(\sum_{t \in T} -\hat{c}_{Pi} \Delta P_{Gi}^t + F_i(1 - z_i) \right) \tag{10.28}$$

Subject to:

$$\Delta P_{Gi}^t + (1 - z_i)P_{Gi}^{max} \leq P_{Gi}^{max} - P_{Gi}^{ot} \quad : v_{Pi}^{+t} \qquad \forall i \in PG, \forall t \in T \tag{10.29}$$

$$-\Delta P_{Gi}^t \leq P_{Gi}^{ot} \qquad\qquad : v_{Pi}^{-t} \qquad \forall i \in PG, \forall t \in T \tag{10.30}$$

$$\sum_{i \in PG} \Delta P_{Gi}^t = 0 \qquad\qquad :\beta_P \qquad \forall t \in T \tag{10.31}$$

$$1 - z_i \leq 1 \qquad\qquad :U_i \tag{10.32}$$

Noting that $0 \leq 1 - z_i \leq 1$ is equivalent to $0 \leq z_i \leq 1$, defining $\Delta P_{Gi}^t = P_{Gi}^t - P_{Gi}^{ot}$ and
repeating the procedure used to simplify the RO formulation primal problem we can re-
express this formulation as:

---

[176]   This allows us to interpret the primal objective function in a similar manner as for the RO model.

$$-\sum_{i\in PG}\sum_{t\in T}\left(c_{Pi}P_{Gi}^{ot}+\frac{f_i}{P_{Gi}^{max}}P_{Gi}^{ot}+F_i\delta_{Fi}\right)+\mathbf{Minimise}_{z^{PG},\,\square\,P_G^{T\cup PG}}\sum_{i\in PG}\sum_{t\in T}\left(c_{Pi}P_{Gi}^{t}+\frac{f_i}{P_{Gi}^{max}}P_{Gi}^{t}+z_iF_i\right)$$

$$(10.33)$$

Subject to:

$$0\le P_{Gi}^{ot}\le z_iP_{Gi}^{max} \qquad\qquad \forall i\in PG,\forall t\in T \quad (10.34)$$

$$\sum_{i\in PG}P_{Gi}^{t}=P_{D}^{t} \qquad\qquad \forall t\in T \qquad\qquad (10.35)$$

$$0\le z_i\le 1 \qquad\qquad\qquad\qquad (10.36)$$

As for the RO model, (10.33) describes the difference in costs between the optimal solution of the relaxed problem and the relaxed costs of the observed solution, and by comparing (10.33) with (10.22) we can conclude that the compensation provided by the MSRO formulation, like that provided for the RO model, equals the compensation required for a linear relaxation of the integer problem, plus an extra component (the Ω_{fi} terms) to account for the difference between the actual average cost of the observed dispatch and the minimum average cost. As for the RO model, the minimum compensation will occur when $P_G^o=P_G^*$, but this compensation and the corresponding best compromise market clearing prices will not be consistent with the optimal integer dispatch unless P_G^* happens to correspond to the optimal integer dispatch.

10.5. Conclusions

In this chapter we have discussed ways in which commitment cost might be recovered. These costs are related to integer primal decision variables, and consequently their recovery is not guaranteed in a marginal cost based framework. Their are two commonly suggested approaches to recovering commitment costs, these being to either impose a charge separate from the spot market transactions or to modify the spot prices. While a price modification approach like Ramsey Pricing has attractive properties, it is difficult to implement in its pure form, and its results might well be approximated by allowing generating companies to exercise some market power. The greedy algorithm of S. Smith (1993) is a good duality based spot price modification procedure, but may not be able to account for enough of the detail of period by period dispatch to be useful in a dispatch based pricing context.

We have proposed a number of variants on these methods, or new methods, to recover commitment costs. A very simple approach proposed was to use a dispatch based pricing model to maximise the transmission networks rents, effectively reducing the need for commitment cost recovery. This is unlikely to make a big impact, however. The use of

uplift was proposed in the context of dispatch based pricing. If applied solely to the pricing reference node the nodal uplift charge will tend to increase with electrical distance from the reference node, leading to disproportionate, and unjustifiable, allocation. Alternatively, if applied to more than one node in a transmission system this uplift will violate the consistency of inter-nodal price differentials.

Best compromise pricing has been proposed as a promising approach for addressing commitment cost recovery. We have developed a best compromise pricing formulation to provide compensation for a single period problem involving running overheads (the RO formulation). This is just the standard best compromise formulation but with the marginal costs of generating units being replaced with the corresponding minimum average costs, and to which a constant compensation term is added to account for the difference between the minimum average cost and the actual observed average cost. This formulation will only produce market clearing prices and compensation consistent with the optimal integer solution if this happens to be the observed solution, otherwise it will produce market clearing prices consistent with the optimal solution to the linear relaxation of the integer problem, though it will provide enough compensation to cover the running overheads.

Applying best compromise pricing to general multiple period integer problems is more difficult. We have suggested a number of simpler formulations which might be used, potentially accepting a higher compensation requirement as a trade-off against problem complexity. The simplest of these approaches involved no explicit mechanism to recover commitment costs, forcing generating companies to account for these in their offers. This could potentially give rise to significant dispatch inefficiencies. Another approach involves determining compensation given the observed dispatch and some estimate of the energy price, without explicitly minimising the compensation. A multiple period version of the RO model, which accounts for start-up costs in addition to running overheads was explored. This MSRO model exhibits the same behaviour as the RO model with compensation equalling the compensation required for a linear relaxation of the integer problem plus some constant compensation terms to recover commitment costs.

While the MSRO model was developed subject to some rather restrictive assumptions, we have demonstrated that best compromise pricing can allow dispatch based prices to be determined which account for commitment costs.

Chapter 11

Conclusions

11.1. Introduction

The application of spot pricing to electricity markets has been proposed for some time, but only in the last 15 years has it become possible for these prices to provide a detailed representation of the economics of power system dispatch. This reflects the fact that spot pricing has become an increasingly popular approach for pricing for transmission network usage, as well as facilitating market transactions between power producers and consumers. The aim of this work has been to extend the theory of electricity spot pricing so as to make the approach more compatible with, and applicable to, actual markets. This has involved increasing the generality and robustness of the theory, while providing a better understanding of the behaviour of prices, and of the forces which drive that behaviour.

11.2. Summary of Results

So as to facilitate an increase in the generality and robustness of spot pricing theory we have proposed, in Chapter 3, a broad philosophical interpretation of spot pricing, which we refer to as Dispatch Based Pricing. Dispatch Based Pricing places the responsibility on the dispatcher to determine prices which satisfy the requirement that:

> *Prices should be determined, after the fact, so as to be consistent with the observed dispatch, in the sense that they provide a rational economic explanation for it, given the price "floors" and "ceilings" expressed by the sellers and buyers in their pre-dispatch offers and bids.*

This approach not only allows pricing variations to be explained by optimally binding constraints, as considered by Caramanis et al. (1982), Schweppe et al. (1988), Hogan (1991, 1992), and Hogan et al. (1995), but allows sub-optimality to be a valid explanation. This allows the dispatcher to control the system in a secure and economically rational way, knowing that the economic implications of that dispatch, and the

complexities of the transmission network, will be consistently reflected in prices which deliver commercial and economic benefits to all parties.

To achieve an economically consistent explanation of an optimal power system dispatch it is necessary to represent as wide a variety of dispatch constraints into the pricing model as possible. Each additional constraint which is modelled provides a new option for explaining the economics of a dispatch, and hence can lead to more accurate pricing. In Chapter 4 we derived a pricing model for an AC power system. Following Hogan (1991), this model is the dual of the linearisation about an observed solution of a, possibly notational, Optimal Power Flow formulation. This means that our model is based on a "snapshot" of the system state. Our Optimal Power Flow formulation employs a PVQ representation of the primal independent variables, which provides a more general representation of the relationship between the electrical variables than the PQ representation used by Hogan (1991) and Hogan et al. (1995). The OPF includes active and reactive power conservation constraints, constraints on dependent reactive power injections and voltage magnitudes, and a general constraint. Given an optimal solution to this OPF the dual objective function is redundant provided that the complementary slackness conditions are imposed directly onto the dual variables, allowing us to substitute an arbitrary pricing objective.

The prices produced by the resulting linear programming problem are driven by the observation of what appears to be out of merit order dispatch, and the physical constraints which explain it, avoiding the need to monitor or reconstruct the dispatch processes themselves. We have described the basic economic interpretation of these prices, and demonstrated that they produce first order incentives consistent with the OPF formulation. While the snapshot approach may appear to abstract somewhat from reality, we have demonstrated, for example, that the primal variables which are constrained and sub-optimal in one time frame may appear optimal and unconstrained in others with different prices resulting in each case. Consequently, we argue, the form of the "correct" pricing model varies with the scope of the dispatch problem, and as the scope of the dispatch problem changes the nearer the dispatcher gets to the actual dispatch period, it is reasonable to price based on the actual snapshot of the dispatch, while incorporating other effects as additions to this basic model.

The general constraint of Chapter 4 was used in Chapter 5 to derive prices for both constraints on transmission flow and transmission line heating. The impact of these constraints for cyclic lines were demonstrated to be somewhat more complicated than for acyclic lines, the latter being the more commonly considered case. We have described the prices around a cyclic line in terms of the "Spring Washer Effect". We show that when a

link in a cyclic network becomes binding that two marginal nodes are required to react simultaneously so as to satisfy this constraint while meeting changes in demand. As these nodes will have different prices, the prices around the network are forced to rise from the side of the constraint nearest the cheapest marginal node around to side nearest the most expensive marginal node.

The general constraint of Chapter 4 was also used to study spinning reserve target constraints in Chapter 6. We modelled constraints which are consistent with the present reserve requirements in New Zealand. These constraints require that, in the event that the largest single source of power in each island fails, the remaining generating units have the ability to increase their output within several seconds to stop the system frequency falling, and then to make up for the lost power production. We have demonstrated that the price for reserve capacity should fully compensate those generating units which provide reserve for the opportunity costs incurred on the energy production foregone. This model ignores the variety of contingencies which may occur, and the way in which reserve may be used to meet them.

The model of Chapter 4 can only explain the examples of out of merit order dispatch which are notified to it by setting price constraints corresponding to constraints on the dispatch. Even when all constraints which are believed to have been binding within the period are accounted for, some inconsistencies are likely to remain. These may be due to problems in interpreting the data, to constraints of a type not accounted for in the model, or which have not been properly communicated to the model, or to genuinely sub-optimal dispatch.

Past approaches to accounting for inconsistencies in dispatches, have taken no account of the economic costs of these inconsistencies, instead relying on arbitrary remedies. In Chapter 7 we presented a best compromise pricing approach, based on a goal programming formulation, which determines spot prices which minimise the (possibly notional) compensation payments that need be made to each market participant so as to account for any discrepancy between their bids and offers and the spot price. When applied to a truly linear problem, best compromise pricing reproduces the optimal spot prices. The approximations required to model more realistic non-linear problems may mean that while no one buying or selling power will make a loss on the transaction that occurs in the sub-optimal solution, the best compromise spot prices will not necessarily match the "true" spot prices (if these exist), and hence not all of the welfare loss will necessarily be recovered by compensation. More research should be conducted in this area.

If a dispatcher is required to pay compensation, then this might be collected via some annual fee. In minimising the payouts from this fund, an incentive is created for the dispatcher to efficiently operate the system. If the dispatcher is the grid owner an incentive may exist for the dispatcher to make a trade-off between claiming that a dispatch is sub-optimal or more tightly constrained than is really the case. This situation can lead to the dispatcher making revenues far in excess of what they should be. To address this situation we have proposed the use of capacity rights, which market participants have an incentive to acquire anyway, so as to transfer such excess rents from the dispatcher back to the market participants.

Our analysis of dispatch based pricing was initially based on the assumption that we were only faced with a single period, deterministic, dispatch problem involving no integer variables. In Chapters 8 to 10 we demonstrate that the dispatch based pricing philosophy has the flexibility to allow these complexities to be addressed within a spot pricing framework. It must be noted, however, that their is potential for significant further work in these areas.

In Chapter 8 we considered three ways of addressing inter-temporal problems when a detailed multi-period representation is considered too complex. An offer price modification approach involved generating companies modifying their offer prices so as to account for the inter-temporal effect. A quantity targeting approach involved having generating companies fix the level at which they wish to generate, effectively relieving the dispatcher of all responsibility to explain this dispatch and placing all the risk on to the generating company. The third approach involved each generating company explicitly informing the dispatcher of any ramping constraints, hence requiring that the dispatcher take responsibility for these constraints, explicitly accounting for them via dispatch based pricing and best compromise pricing. What ever approach is used, effectively the same judgements must be made, and it is moot whether it is more efficient or effective for these to be made by a central dispatcher or by many independent market participants. In Chapter 8 we also conclude that second order incentive may exist for a profit maximising generating company to modify the form of its ramping constraints, and note that while the offer modification approach may go some way to limiting this effect.

The stochastic reserve problem was discussed in Chapter 9. Given the complexity of accurately modelling this problem we have suggested that dispatch based prices be based on a simpler "Optimal Reserve Targeting" (ORT) model. This model, which is similar to one proposed by Caramanis et al. (1987), effectively determines the optimal level of reserve to be carried without blackout being called upon. For this model we have shown that dispatch based prices based on intentions should be consistent with prices based on

dispatch outcomes, even though prices do not necessarily rise in contingency states relative to the base state. The base state price does not necessarily correspond to the marginal cost of any generating source, but accounts for the price difference between the marginal cost of meeting an incremental increase in demand, and the expected marginal cost of calling on reserve in a contingency state. The incentives created by the ORT model prices have been analysed and used to describe the factors which influence how much reserve each source provides. We have not considered the inter-temporal aspects associated with the stochastic reserve problem.

The recovery of commitment costs, these being related to primal integer variables, was discussed in Chapter 10. While a range of methods have been proposed in the literature, none of these adequately address the short run commitment cost recovery problem to the same level of detail proposed for marginal cost recovery. While some commitment costs may be recovered by maximising spot market rents or by "uplifting" the price at the spot pricing reference node these methods are unlikely to make a significant contribution. We therefore proposed the use of best compromise pricing to provide compensation to cover those commitment costs not recovered via conventional spot prices. We have demonstrated that a best compromise pricing model based on the linear relaxation of an integer dispatch problem (with some extra compensation terms, dependent only the observed dispatch, added to the best compromise pricing objective function) can adequately recover commitment costs for single period problems and special cases of the general multiple period problem.

References

R. H. Ansell and R. J. Lermit (1983), An Optimal Pricing Structure for Electricity, *Proceedings of the Operational Research Society of New Zealand*, pages 57-67.

L. T. Anstine, R. E. Burke, J. E. Casey, R. Holgate, R. S. John, H. G. Stewart (1963), Application of Probability Methods to the Determination of Spinning Reserve Requirements for the Pennsylvania-New Jersey-Maryland Interconnection, *IEEE Transactions of Power Apparatus and Systems*, Volume 82, pages 726-735.

R. Bacher (1992), Power System Models, Objectives and Constraints in Optimal Power Flow Calculations. In K. Frauendorfer, H. Glavitsch, and R. Bacher (editors), *Optimization in Planning and Operation of Electric Power Systems,* Schweizerishe Vereinigung Für Operations Research and Association Suisse De Recherche Opérationnelle, Zurich. Pages 159-198, October.

R. Baldick (1995), The Generalised Unit Commitment Problem. *IEEE Transactions on Power Systems*, 10(1), pages 465-475, February.

R. Baldick and E. Kahn (1995), Contract Paths, Phase-Shifters, and Efficient Electricity Trade. Working paper PWP-035, Program on Workable Energy Regulation, University of California Energy Institute, Berkeley, October.

M. L. Balinsky and W. J. Baumol (1968), The Dual in Nonlinear Programming and its Economic Interpretation, *The Review of Economic Studies*, Vol. 35, No. 3, pages 237-256, July.

M. L. Baughman and S. N. Siddiqi (1991), Real-time Pricing of Reactive Power: Theory and Case Study Results. *IEEE Transactions on Power Systems*, 6(1), pages 23-29, February.

M. L. Baughman, S. N. Siddiqi, and J. Zarnikau (1992), Advanced Pricing in Electrical Systems. Discussion Draft, Department of Electrical and Computer Engineering and Center for Energy Studies, University of Texas at Austin, October.

W. J. Baumol and D. F. Bradford (1970), Optimal Departures from Marginal Cost Pricing, *American Economic Review*, pages 265-283, June.

M. S. Bazaraa and J. J. Jarvis (1977), *Linear Programming and Network Flows*. John Wiley and Sons, New York.

G. Bergendahl (1983), Marginal Cost Pricing and Revenue Reconciliation for Electricity, *European Journal of Operations Research*, Vol. 13, pages 274-277.

A. W. Berger and F. C. Schweppe (1989), Real Time Pricing to Assist in Load Frequency Control. *IEEE Transactions on Power Systems*, 4(3), pages 920-926, August.

S. Bernstein (1988), Competition, Marginal Cost Tariffs and Spot Pricing in the Chilean Electric Power Sector. *Energy Policy*, 4(16), pages 369-377, August.

S. Bernstein and R. Agurto (1993), Transmission Pricing in the Power Sector in Chile. SYNEX Consulting Engineers, Santiago, Chile, May.

R. E. Bohn, M. C. Caramanis, and F. C. Schweppe (1984), Optimal Pricing in Electrical Networks Over Space and Time. *Rand Journal of Economics*, 15(3), pages 360-376.

R. E. Bohn, B. W. Golub, R. D. Tabors, and F. C. Schweppe (1984), Deregulating the Generation of Electricity Through the Creation of Spot Markets for Bulk Power. *The Energy Journal*, 5(2), pages 71-91.

J. F. Boshier (1974), Nonlinear Programming Applied to Power System Voltage Control. *New Zealand Operational Research*, 2(1), pages 40-48, January.

J. Bushnell and S. Stoft (1995), Transmission and Generation Investment In A Competitive Electric Power Industry. Program on Workable Energy Regulation, PWP-030, May.

California Public Utilities Commission (1995), Proposed Policy Decision Adapting a Preferred Industry Structure, Docket R. 94-04-031, May 24.

M. C. Caramanis (1982), Investment Decisions and Long-Term Planning Under Electricity Spot Pricing. *IEEE Transactions on Power Apparatus and Systems*, PAS-101(12), pages 4640-4648, December.

M. C. Caramanis, R. E. Bohn and F. C. Schweppe (1982), Optimal Spot Pricing: Practice and Theory. *IEEE Transactions on Power Apparatus and Systems*, PAS-101(9), pages 3234-3245, September.

M. C. Caramanis, R. E. Bohn, and F. C. Schweppe (1987), System Security Control and Optimal Pricing of Electricity. *Electrical Power and Energy Systems*, 9(4), pages 217-224, October.

M. C. Caramanis, N. Roukos, and F. C. Schweppe (1989), WRATES: A Tool for Evaluating the Marginal Cost of Wheeling, IEEE Transactions on Power Systems, 4(2), May.

H.-P. Chao, S. S. Oren, S. A. Smith, and R. B. Wilson (1986), Multilevel Demand Subscription Pricing for Electric Power, *Energy Economics*, 8(4), pages 199-217, October.

H.-P. Chao and S. P. Peck (1995), Market Mechanisms for Electric Power Transmission. Presented at "Energy Model for Policy and Planning", First Joint International Symposium, London Business School and IFORS.

H.-P. Chao and R. Wilson (1987), Priority Service: Pricing, Investment, and Market Organization. *The American Economic Review*, 77(5), pages 899-916.

D. Chattopadhyay, K. Bhattacharya, J. Parikh (1995), Optimal Reactive Power Planning and its Spot Pricing: An Integrated Approach. Presented at the IEEE Power Engineering Society Winter Meeting, New York.

R. H. Coase (1970), The Theory of Public Utility Pricing and its Application. *Bell Journal of Economics and Management Science*, 1(1), pages 113-128.

M. A. Crew and P. R. Kleindorfer (1979), *Public Utility Economics*. St Martins Press, New York.

J. G. Culy, E. G. Read, and B. D. Wright (1995), The Evolution of New Zealand's Electricity Supply Structure. To appear in R. Gilbert and E. Kahn (editors), *International Regulation of Electric Power*, Cambridge University Press.

N. H. Dandachi, M. J. Rawlins, O. Alsac, M. Prais, and B. Stott (1995), OPF for Reactive Pricing Studies on the NGC System. Presented at IEEE PES PICA Conference, Salt Lake City, May.

A. P. Della Valle (1988), Short-Run Versus Long-Run Marginal Cost Pricing. *Energy Economics*, pages 283-286, October.

L. G. Dias and M. E. El-Haway (1990), Voltage Effects in Optimal Power Flow Including Load Modelling: A Comparison of Fixed Versus Floating Generator

Voltages. *Proceedings of the Tenth Power Systems Computation Conference*, Graz, Austria, August.

G. R. Drayton, B. W. Lamar, E. G. Read, and A. Turner (1992), Optimization of Generation and Spinning Reserve in an Electrical Power System *Proceedings of the Operational Research Society of New Zealand*, Christchurch, pages 142-149, August.

G. R. Drayton, E. G. Read, B. J. Ring and M. C. Rosevear (1993), Transmission System Pricing Model: Record of Experimental Tests. Energy Modelling Research Group, Department of Management, University of Canterbury. Draft report.

M. A. Einhorn (editor) (1994), *From Regulation to Competition: New Frontiers in Electricity Markets*. Kluwer Academic Press, Boston.

Electric Power Research Institute (EPRI) (1995), A Primer on Electric Power Flow for Economists and Utility Planners. EPRI TR-104604s, February.

T. W. Gedra (1992), Calls and Callable Forwards for Electric Power Markets. In the Proceedings of the NSF workshop "The Impact of a Less-Regulated Utility Environment on Power System Control and Security", pages 3-14.

H. Glavitsh (1992), Use of Linear and Quadratic Programming Techniques in Exploiting the Non-linear Features of the Optimal Power Flow. In K. Frauendorfer, H. Glavitsh, and R. Bacher (editors), *Optimisation in Planning and Operation of Electric Power Systems*, Schweizerishe Vereinigung Für Operations Research and Association Suisse De Recherche Operationnelle, Zurich. Pages 199-235, October.

R. J. Green (1994), The British Electricity Spot Market. Proceedings of the 17th Annual International Conference of the International Association of Energy Economics, Stravanger, Norway, May 1994.

R. J. Green and D. M Newbery (1992), Competition in the British Electricity Spot Market. *Journal of Political Economy*, 100(5), pages 929-953.

K. A. Heinz (1995), Argentina's Wholesale Electric Market: Proposals for Increasing Market Efficiency. Southern Electric International, Inc.

A. Henney (1994), *A Study of the Privatisation of the Electricity Supply Industry in England and Wales*. EEE Ltd, London.

L. Hjalmarsson (1992), The Market for Electricity in Scandinavia: Regulation and Performance. To appear in R. Gilbert and E. Kahn (editors), *International Regulation of Electric Power*, Cambridge University Press.

W. W. Hogan (1991), Contract Networks for Electric Power Transmission: Technical Reference. Energy and Environmental Policy Center, John F. Kennedy School of Government, Harvard University, December.

W. W. Hogan (1992), Contract Networks for Electric Power Transmission *Journal of Regulatory Economics*, 4, pages 211-242.

W. W. Hogan (1993), Markets in Real Electricity Networks Require Reactive Prices. *The Energy Journal*, 4(3), pages 171-200.

W. W. Hogan, E. G. Read, and B. J. Ring (1995), Using Mathematical Programming for Electricity Spot Pricing. Submitted to *International Transactions in Operational Research*.

J. A. Holtan, I. Wangensteen, and K. Livik (1994), Deregulation of Electricity Supply: The Norwegian Experience. *Proceedings of the 17th Annual International Conference of the International Association of Energy Economics*, Stravanger, Norway, May.

M. Huneault and F. D. Galiana (1991), A Survey of the Optimal Power Flow Literature. *IEEE Transactions on Power Systems*, 6(2), pages 762-770.

P. L. Joskow (1976), Contributions to the Theory of Marginal Cost Pricing. *The Bell Journal of Economics*, 7(1), pages 197-206.

P. L. Joskow and R. Schmalensee (1983), *Markets for Power: An Analysis of Electrical Utility Deregulation*. MIT Press, Cambridge, Massachusetts.

E. Kahn and R. Baldick (1994), Reactive Power is a Cheap Constraint. *The Energy Journal*, 15(4), pages 191-204.

M. Kennedy (1994), Creating a Competitive Electricity Market in New Zealand. *Proceedings of the 17th Annual International Conference of the International Association of Energy Economics*, Stravanger, Norway, May 1994.

S. M. Lee (1972), Goal Programming for Decision Analysis. Auerbach, Philadelphia.

S. C. Littlechild (1991), Spot Pricing of Electricity: An Appraisal of the MIT Model. Originally presented at a Conference in Honour of A. Charnes, Austin, Texas, October 1987 (Revised).

S. C. Littlechild (1994), Recent Developments in Competition and Regulation in Electricity Markets in the United Kingdom. Presentation at Harvard University.

R. Miller and A. Turner (1991), A Spreadsheet Model for Hydro-Thermal-HVDC-Reserve Scheduling. Trans Power (NZ) Ltd, Wellington.

B. M. Mitchell, W. G. Manning, and J. P. Acton (1978), *Peak-Load Pricing*. Ballinger Publishing Company, Cambridge, Massachusetts.

J. A. Muckstadt and S. A. Koenig (1977), An Application of Lagrangian Relaxation to Scheduling in Power-Generation Systems. *Operations Research*, 25(3), pages 387-403.

L. Murphy, R. J. Kaye, and F. F. Wu (1994), Distributed Spot Pricing in Radial Distribution Systems. *IEEE Transactions on Power Systems*, 9(1), 311-317, February.

J. R. Nelson (editor) (1964), *Marginal Cost Pricing in Practice*. Prentice-Hall Series in Management. Prentice-Hall Inc, New Jersey.

D. M. Newberry (1995), Power Markets and Market Power. *The Energy Journal*, 16(3), pages 39-66.

S. S. Oren and S. A. Smith (1992), Design and Management of Curtailable Electricity Service to Reduce Annual Peaks, *Operations Research*, 40(2), pages 213-228, March-April

S. S. Oren, P. T. Spiller, P. Varaiya, and F. Wu (1995), Nodal Prices and Transmission Rights: A Critical Appraisal. *The Electricity Journal*, pages 24-35, April.

H. Outhred (1994), Review of Proposals for Dealing with Network Issues in A Competitive Electricity Industry. Department of Electric Power Engineering Report DEPE940727, University of New South Wales, August

H. Outhred, C. H. Bannistar, R. J. Kaye, Y. B. Lee, D. Sutanto, and R. Manimaran (1988), Electricity Pricing: Optimal Operation and Investment By Industrial Customers. *Energy Policy*, 16(4), pages 384-393, August

H. Outhred and J. Kaye (1994), Incorporating Network Effects in a Competitive Electricity Industry: An Australian Perspective. Department of Electric Power Engineering Report DEPE940607, University of New South Wales, July (Presented to The Institute of Management Sciences Conference, Alaska, June 1994)

E. G. Read (1979), Optimal Operation of Power Systems. Ph.D. Thesis, Department of Operations Research, University of Canterbury.

E. G. Read (1989), Pricing of Transmission Services: Long Run Aspects. In A. J. Turner (editor), *Principles for Pricing Electricity Transmission*. Trans Power (NZ) Ltd, Wellington.

E. G. Read (1991),Transmission Pricing Issues in New Zealand: Status Report. Report to the Electricity Corporation of New Zealand Ltd.

E. G. Read (1992), Short Run Transmission Pricing Issues in New Zealand. Proceedings of the Fifteenth Annual Conference of the International Association for Energy Economics, Tours, France, pages A17-A24.

E. G. Read (1995), OR Modelling for a Deregulated Electricity Sector. Submitted to *International Transactions in Operational Research*.

E. G. Read and B. J. Ring (1995a), Dispatch Based Pricing: Philosophy and Methodology. In A. J. Turner (editor), *Dispatch Based Pricing*. Trans Power (NZ) Ltd, Wellington.

E. G. Read and B. J. Ring (1995b), Dispatch Based Pricing: Theory and Application. In A. J. Turner (editor), *Dispatch Based Pricing*. Trans Power (NZ) Ltd, Wellington.

E. G. Read and B. J. Ring (1995c), Dispatch Based Pricing: Behaviour of Nodal Power Prices. In A. J. Turner (editor), *Dispatch Based Pricing*. Trans Power (NZ) Ltd, Wellington.

E. G. Read and B. J. Ring (1995d), Dispatch Based Electricity Spot Pricing. In R. Siddiqi and M. Einhorn (editors), *Issues in Transmission Pricing and Technology: A Collection of Readings*, Kluwer Academic Press, Boston. Pages 183-206.

E. G. Read, B. J. Ring, and M. C. Rosevear (1995), Dispatch Based Pricing: Technical Reference. In A. J. Turner (editor), *Dispatch Based Pricing*. Trans Power (NZ) Ltd, Wellington.

E. G. Read and D. P. M. Sell (1989), Pricing and Operation of Transmission Services, Report 1: Short Run Aspects. In A. J. Turner (editor), *Principles for Pricing Electricity Transmission*, Reports from Arthur Young and E. G. Read. Trans Power (NZ) Ltd, Wellington.

B. J. Ring (1995), A General Stochastic Reserve Pricing Model, Discussion Draft, Energy Modelling Research Group, Department of Management, University of Canterbury.

B. J. Ring and E. G. Read (1994c), Pricing for Reserve Capacity in a Competitive Electricity Market. Presented to *TIMS XXXII,* Anchorage, Alaska, June.

B. J. Ring and E. G. Read (1995), Short Run Pricing in Competitive Electricity Markets, to appear in the *Canadian Journal of Economics*.

B. J. Ring, E. G. Read, and G. R. Drayton (1993a), Optimal Pricing for Reserve Electricity Generation Capacity. *Proceedings of the Operational Research Society of New Zealand*, Auckland, pages 84-91, August.

B. J. Ring, E. G. Read, and G. R. Drayton (1993b), An Analytical Framework for Pricing Spinning Reserve. Report to Trans Power (NZ) Ltd.

M. Rivier and I. J. Pérez-Arriaga (1991), Cost Allocation of Transmission Services. Presented at the 1991 Power Industry Computer Applications Conference, Baltimore, May.

M. C. Rosevear, E. G. Read, and N. R. Watson (1992), Loss Differential System Charging for Reactive Power. Report to Trans Power (NZ) Ltd.

M. C. Rosevear and B. J. Ring (1992), Implementation of Transmission Pricing Using a GAMS Model. Proceedings of the Operational Research Society of New Zealand, Christchurch, pages 159-165, August.

L. E. Ruff (1992), Competitive Electricity Markets: Economic Logic and Practical Implementation. Presented at the Fifteenth Annual Conference of the International Association for Energy Economics, Tours, France.

T. J. Scott and E. G. Read (1992), Optimising Reservoir Management in a Deregulated Electricity Market, Proceedings of the Operational Research Society of New Zealand, Auckland, pages 92-99, August.

T. J. Scott and E. G. Read (1995), Modelling Hydro Generator Gaming in a Deregulated Electricity Market. To appear in *International Transactions in Operational Research*.

F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn (1988), *Spot Pricing of Electricity*. Kluwer Academic Publishers, Norwell.

F. C. Schweppe, R. D. Tabors, H. Kirtley, H. Outhred, F. Pickel and A. Cox (1980), Homeostatic Utility Control. *IEEE Transactions on Power Apparatus and Systems*, PAS-99(3), pages 1151-1163, May-June.

S. N. Siddiqi and M. L. Baughman (1992), Reliability Differentiated Real-time Pricing of Electricity. Paper 92WM115-6PWRS, IEEE 1992 Winter Power Meeting, New York, pages 26-30, January.

S. N. Siddiqi and M. L. Baughman (1994), Optimal Pricing of Non-Utility Generated Electric Power. *IEEE Transactions on Power Systems*, 9(1), pages 397-403, February.

S. A. Smith (1993), A Linear Programming Model for Real-Time Pricing of Electric Power Service. *Operations Research*, 4(3), pages 470-483, May-June.

V. L. Smith (1993), Can Electric Power - A "Natural Monopoly" - Be Deregulated? In H. H. Landsberg (editor), *Making National Energy Policy*, Resources for the Future Inc., Washington D. C.

P. T. Spiller and L. V. Martorell (1992), How Should It Be Done? Electricity Regulation in Argentina, Brazil, Uruguay and Chile. To appear in R. Gilbert and E. Kahn (editors), *International Regulation of Electric Power*, Cambridge University Press.

W. O. Stadlin (1971), Economic Allocation of Regulating Margin. *IEEE Transaction on Power Apparatus and Systems*, PAS-90, pages 1776-1781.

C. G. Stalon (1987), Wheeling, Competition, and Economic Efficiency. *Energy Systems and Policy*, 11, pages 77-83.

B. Stott, O. Alsac, and A. J. Monticelli (1987), Security Analysis and Optimization. *Proceedings of the IEEE*, 75(12), pages 1623-1644, December.

R. D. Tabors (1994), Transmission System Management and Pricing: New Paradigms and International Comparisons. *IEEE Transactions on Power Systems*, 9(1), pages 206-215, February.

R. D. Tabors and M. C. Caramanis (1994), Transmission Pricing: Can it Be Done in Real Time. *Proceedings 1994 Innovative Electricity Pricing*. EPRI TR-103639, pages 207-220, February.

R. D. Tabors, F. C. Schweppe, and M. C. Caramanis (1989), Utility Experience with Real Time Rates. *IEEE Transactions on Power Systems*, 4(2), pages 463-471, May.

B. Tenenbaum, R. Lock and J. Barker (1992), Electricity Privatization. *Energy Policy*, 20(12), pages 1134-1160, December.

B. Tenenbaum (1993), Spot Pricing Markets in the United States. Lecture notes.

U.S. Congress, Office of Technology Assessment (1989), Electric Power Wheeling and Dealing: Technological Considerations for Increasing Competition. Report OTA-E-490, US Government Pricing Office, Washington D.C., May.

S. V. Vankatesh, W.-H. E. Liu, and A. D. Papalexopoulos (1992), Least Squares Solution for Optimal Power Flow Sensitivity Calculation. *IEEE Transactions on Power Systems*, 7(3), pages 1394-1401, August.

W. S. Vickrey (1979), Responsive Pricing Based on Marginal Cost as a Means of Promoting Efficient Energy Use. In J. Dunkerly (editor), *International Energy Strategies: Proceedings of the 1979 IAEE/RFF Conference*, Washington D. C., June, Oegeschlager, Gunn and Hain Publishers, Inc.

M. Wilkinson (1989), Power Monopolies and the Challenge of the Market: American Theory and British Practice. Report E-89-12, Energy and Environmental Policy Center, John F. Kennedy School of Government, Harvard University, September.

P. Wolfe (1961), A Duality Theorem for Non-linear Programming, *Quarterly of Applied Mathematics*, 19(3), pages 239-244.

A. J. Wood and B. F. Wollenberg (1984), *Power Generation, Control, and Operation*. John Wiley and Sons, New York

F. Wu, P. Varaiya, P. Spiller, and S. Oren (1994), Folk Theorems on Transmission Access: Proofs and Counter Examples. Working paper PWP-023, Program on Workable Energy Regulation, University of California Energy Institute, Berkeley, October.

Appendix A

Derivation of a Dispatch Based Pricing Model

A.1. Introduction

In this Appendix the dual pricing equations (4.15) to (4.24) are derived from the primal OPF equations (4.1) to (4.14). The prices we wish to derive each correspond to the change in the cost of operating the power system when the availability of one of the resources is incrementally changed. Consequently we can form these prices by linearising the OPF in the immediate vicinity of an observed dispatch. The mathematical programming dual of this linearisation determines the prices required. This technique is essentially that used by Hogan (1991, 1992) but with some modifications.

A.2. Linearisation of the Primal OPF

In this section we describe the steps involved in linearising the OPF, presenting the resulting linear problem at the end.

A.2.1. Linearising Non-Differentiable Functions

While most of the functions in the OPF described by (4.1) to (4.14) are assumed to be convex[177] with unique derivatives, the cost function (4.1) and general constraints (4.6) are assumed to include functions which are convex but have points at which they are non-differentiable with respect to some variables[178]. That is, these functions may have "corners" which imply different derivatives on either "side" of an observed solution. Consequently, given such a function, $F(y)$, we must linearise it as:

---

[177]   See Section 4.2.4.

[178]   The general constraints are the only constraints which can involve derivatives with respect to dependent variables, though.

$$F(y) \approx F(y^*) + \frac{\partial F}{\partial y^-} y^- + \frac{\partial F}{\partial y^+} y^+ \tag{A.1}$$

$$y^+, y^- \geq 0 \tag{A.2}$$

Where y^* is the observed value of variable y. The values of the left and right hand sides of (A.1) are, by definition, the same at the observed solution. It should be noted that an increase in y^- corresponds to a decrease in y. This convention differs to that used by Hogan who had an equivalent variable to y^- which increased with y but was constrained to be less than y^*. While our convention produces the same results as Hogan's, it seems a more intuitive way of describing incremental deviations from an observed solution. Provided that the function $F(y)$ is convex[179], we will always have:

$$\frac{\partial F}{\partial y^+} > -\frac{\partial F}{\partial y^-} \tag{A.3}$$

This linearisation and (A.3) are demonstrated by Figure A.1.
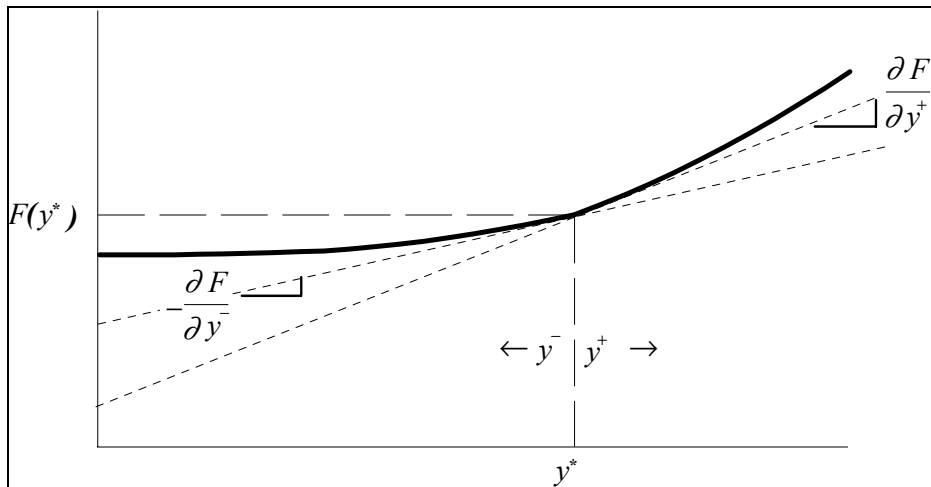
Figure A.1: Linearisation of a discontinuous function

The variables which are linearised in this manner are:

$$P_{Gi} = P_{Gi}^* + P_{Gi}^+ - P_{Gi}^- \qquad \forall i \in PXS \tag{A.4}$$

$$Q_{Gi} = Q_{Gi}^* + Q_{Gi}^+ - Q_{Gi}^- \qquad \forall i \in PXS \tag{A.5}$$

$$x_\ell = x_\ell^* + x_\ell^+ - x_\ell^- \qquad \forall \ell \in ND \tag{A.6}$$

We simplify the linearised objective function by replacing the cost derivative terms with equivalent marginal cost terms. That is:

---

[179]   Non-convex functions can only be handled by using integer variables. These introduce a range of issues which may be significant, but which cannot be handled by the kind of dual formulation used here. See Chapter 10 for further discussion of this.

$$c_{Pi}^{+} = \frac{\partial C}{\partial P_{Gi}^{+}} \qquad\qquad i \in PXS \qquad (A.7)$$

$$c_{Pi}^{-} = -\frac{\partial C}{\partial P_{Gi}^{-}} \qquad\qquad i \in PXS \qquad (A.8)$$

$$c_{Qi}^{+} = \frac{\partial C}{\partial Q_{Gi}^{+}} \qquad\qquad i \in PXS \qquad (A.9)$$

$$c_{Qi}^{-} = -\frac{\partial C}{\partial Q_{Gi}^{-}} \qquad\qquad i \in PXS \qquad (A.10)$$

$$c_{x\ell}^{+} = \frac{\partial C}{\partial x_{\ell}^{+}} \qquad\qquad \ell \in ND \qquad (A.11)$$

$$c_{x\ell}^{-} = -\frac{\partial C}{\partial x_{\ell}^{-}} \qquad\qquad \ell \in ND \qquad (A.12)$$

With these definitions the linearised objective function, with constants ignored, can be represented as:

$$\sum_{i \in PXS}\left(c_{Pi}^{+}P_{Gi}^{+} - c_{Pi}^{-}P_{Gi}^{-}\right) + \sum_{i \in PXS}\left(c_{Qi}^{+}Q_{Gi}^{+} - c_{Qi}^{-}Q_{Gi}^{-}\right) + \sum_{\ell \in ND}\left(c_{x\ell}^{+}x_{\ell}^{+} - c_{x\ell}^{-}x_{\ell}^{-}\right)$$

A.2.2. Linearisation of Differentiable Functions

Most of the OPF constraints can be linearised using a standard Taylor expansion of the form:

$$F(y) \approx F(y^{*}) + \frac{\partial F}{\partial y}(y - y^{*})$$

$$= F^{*} + \frac{\partial F}{\partial y}(y - y^{*})$$

To demonstrate this for one of the OPF constraints we consider the active power energy conservation constraint:

$$\sum_{i \in PXS}\left(P_{Gi} - P_{Di}\right) - L_{P}^{A}\left(\boldsymbol{P_{G}^{PX}} - \boldsymbol{P_{D}^{PX}}, \boldsymbol{Q_{G}^{PQ}} - \boldsymbol{Q_{D}^{PQ}}, \boldsymbol{V^{PVS}}\right) = 0$$

Linearising the loss function, ignoring second and higher order terms, and substituting from (A.4) and (A.5) gives:

$$\sum_{i \in PXS}\left(P_{Gi}^{*} + P_{Gi}^{+} - P_{Gi}^{-} - P_{Di}\right) - L_{P}^{*} - \sum_{i \in PX}\frac{\partial L_{P}^{A}}{\partial^{A}P_{i}}\left(\left(P_{Gi}^{*} + P_{Gi}^{+} - P_{Gi}^{-} - P_{Di}\right) - \left(P_{Gi}^{*} - P_{Di}^{*}\right)\right)$$

$$- \sum_{i \in PQ}\frac{\partial L_{P}^{A}}{\partial^{A}Q_{i}}\left(\left(Q_{Gi}^{*} + Q_{Gi}^{+} - Q_{Gi}^{-} - Q_{Di}\right) - \left(Q_{Gi}^{*} - Q_{Di}^{*}\right)\right) - \sum_{i \in PVS}\frac{\partial L_{P}^{A}}{\partial^{A}V_{i}}\left(V_{i} - V_{i}^{*}\right) = 0$$

The power derivatives are represented with respect to net nodal power injections, as an increase in demand has an identical marginal impact to a decrease in generation[180]. By shifting all constant terms to the right hand side and cancelling like terms, this can be re-expressed in its final form as:

$$\sum_{i \in PXS}\left(P_{Gi}^{+}-P_{Gi}^{-}-P_{Di}\right)-\sum_{i \in PX}\frac{\partial L_{P}^{A}}{\partial^{A}P_{i}}\left(P_{Gi}^{+}-P_{Gi}^{-}-P_{Di}\right)-\sum_{i \in PQ}\frac{\partial L_{P}^{A}}{\partial^{A}Q_{i}}\left(Q_{Gi}^{+}-Q_{Gi}^{-}-Q_{Di}\right)-\sum_{i \in PVS}\frac{\partial L_{P}^{A}}{\partial^{A}V_{i}}V_{i}$$

$$= L_{P}^{*}+\sum_{i \in PX}\frac{\partial L_{P}^{A}}{\partial^{A}P_{i}}P_{Di}^{*}+\sum_{i \in PQ}\frac{\partial L_{P}^{A}}{\partial^{A}Q_{i}}Q_{Di}^{*}-\sum_{i \in PVS}\frac{\partial L_{P}^{A}}{\partial^{A}V_{i}}V_{i}^{*}-\sum_{i \in PXS}P_{Gi}^{*}$$

This linearisation differs from Hogan's because of our use of the *PVQ* representation and because of the different linearisation convention used for non-differentiable functions.

A.2.3. Treatment of the Bounds

While it would be possible to substitute (A.4) and (A.5) directly into the primal generating unit bounds we prefer to place separate bounds on the incremental increases and decreases in generation as shown:

$$-P_{Gi}^{+} \geq -P_{Gi}^{max} + P_{Gi}^{*} \qquad\qquad \forall i \in PXS$$

$$-P_{Gi}^{-} \geq P_{Gi}^{min} - P_{Gi}^{*} \qquad\qquad \forall i \in PXS$$

$$-Q_{Gi}^{+} \geq -Q_{Gi}^{max} + Q_{Gi}^{*} \qquad\qquad \forall i \in PXS$$

$$-Q_{Gi}^{-} \geq Q_{Gi}^{min} - Q_{Gi}^{*} \qquad\qquad \forall i \in PXS$$

This form implies that the same cost for an increase (decreases) in generation is assumed for all feasible increases (decreases).

A.2.4   The Final Form of the Linearisation

Applying the methods described in this section to the OPF of (4.1) to (4.14) gives the linearised form shown in (A.13) to (A.26). The dual variables corresponding to each of the primal constraints are indicated. Despite the complexity of the notation, this linearisation transforms the OPF from a non-linear optimisation problem into a much simpler linear programming problem defined about an observed solution. The objective function contains only linearised variables, while the constraints have linearised variables on the left hand side and constant terms on the right.

---

[180]   This is not necessarily the case for general constraints.

$$
\underset{\substack{P_G^{+PXS},P_G^{-PXS},Q_G^{+PXS},Q_G^{-PXS},x^{+NDI},x^{-ND} \geq 0 \\ P_D^{PXS},Q_D^{PXS},V^{PXS}}}{\textbf{Minimise}} \quad \sum_{i \in PXS}\left(c_{Pi}^{+}P_{Gi}^{+} - c_{Pi}^{-}P_{Gi}^{-}\right) + \sum_{i \in PXS}\left(c_{Qi}^{+}Q_{Gi}^{+} - c_{Qi}^{-}Q_{Gi}^{-}\right) + \sum_{\ell \in ND}\left(c_{x\ell}^{+}x_{\ell}^{+} - c_{x\ell}^{-}x_{\ell}^{-}\right) \tag{A.13}
$$

subject to:

Conservation of Power

$$
\sum_{i \in PXS}\left(P_{Gi}^{+} - P_{Gi}^{-} - P_{Di}\right) - \sum_{i \in PX}\frac{\partial L_P^A}{\partial^A P_i}\left(P_{Gi}^{+} - P_{Gi}^{-} - P_{Di}\right) - \sum_{i \in PQ}\frac{\partial L_P^A}{\partial^A Q_i}\left(Q_{Gi}^{+} - Q_{Gi}^{-} - Q_{Di}\right) - \sum_{i \in PVS}\frac{\partial L_P^A}{\partial^A V_i}V_i
$$
$$
= L_P^{*} + \sum_{i \in PX}\frac{\partial L_P^A}{\partial^A P_i}P_{Di}^{*} + \sum_{i \in PQ}\frac{\partial L_P^A}{\partial^A Q_i}Q_{Di}^{*} - \sum_{i \in PVS}\frac{\partial L_P^A}{\partial^A V_i}V_i^{*} - \sum_{i \in PXS}P_{Gi}^{*} \qquad :\lambda_P \tag{A.14}
$$

$$
\sum_{i \in PXS}\left(Q_{Gi}^{+} - Q_{Gi}^{-} - Q_{Di}\right) - \sum_{i \in PX}\frac{\partial L_Q^A}{\partial^A P_i}\left(P_{Gi}^{+} - P_{Gi}^{-} - P_{Di}\right) - \sum_{i \in PQ}\frac{\partial L_Q^A}{\partial^A Q_i}\left(Q_{Gi}^{+} - Q_{Gi}^{-} - Q_{Di}\right) - \sum_{i \in PVS}\frac{\partial L_Q^A}{\partial^A V_i}V_i
$$
$$
= L_Q^{*} + \sum_{i \in PX}\frac{\partial L_Q^A}{\partial^A P_i}P_{Di}^{*} + \sum_{i \in PQ}\frac{\partial L_Q^A}{\partial^A Q_i}Q_{Di}^{*} - \sum_{i \in PVS}\frac{\partial L_Q^A}{\partial^A V_i}V_i^{*} - \sum_{i \in PXS}Q_{Gi}^{*} \qquad :\lambda_Q \tag{A.15}
$$

Dependent Reactive Power Injection at *PV* Nodes

$$
-\sum_{i \in PX}\frac{\partial Q_n^A}{\partial^A P_i}\left(P_{Gi}^{+} - P_{Gi}^{-} - P_{Di}\right) - \sum_{i \in PQ}\frac{\partial Q_n^A}{\partial^A Q_i}\left(Q_{Gi}^{+} - Q_{Gi}^{-} - Q_{Di}\right) - \sum_{i \in PVS}\frac{\partial Q_n^A}{\partial^A V_i}V_i + \left(Q_{Gn}^{+} - Q_{Gn}^{-} - Q_{Dn}\right)
$$
$$
= -Q_{Dn}^{*} + \sum_{i \in PX}\frac{\partial Q_n^A}{\partial^A P_i}P_{Di}^{*} + \sum_{i \in PQ}\frac{\partial Q_n^A}{\partial^A Q_i}Q_{Di}^{*} - \sum_{i \in PVS}\frac{\partial Q_n^A}{\partial^A V_i}V_i^{*} \qquad :\mu_{Qn} \qquad \forall n \in PV \tag{A.16}
$$

Constraint set continues on following page

Dependent Voltage Magnitude at *PQ* Nodes

$$-\sum_{i\in PX}\frac{\partial V_n^A}{\partial^A P_i}\left(P_{Gi}^+ - P_{Gi}^- - P_{Di}\right) - \sum_{i\in PQ}\frac{\partial V_n^A}{\partial^A Q_i}\left(Q_{Gi}^+ - Q_{Gi}^- - Q_{Di}\right) - \sum_{i\in PVS}\frac{\partial V_n^A}{\partial^A V_i}V_i + V_n$$

$$= V_n^* + \sum_{i\in PX}\frac{\partial V_n^A}{\partial^A P_i}P_{Di}^* + \sum_{i\in PQ}\frac{\partial V_n^A}{\partial^A Q_i}Q_{Di}^* - \sum_{i\in PVS}\frac{\partial V_n^A}{\partial^A V_i}V_i^* \qquad :\mu_{Vn} \qquad \forall n\in PQ \quad (A.17)$$

Generalised Constraints

$$\sum_{i\in PXS}\left(\frac{\partial H_h^A}{\partial^A P_{Gi}^+}P_{Gi}^+ + \frac{\partial H_h^A}{\partial^A P_{Gi}^-}P_{Gi}^- + \frac{\partial H_h^A}{\partial^A P_{Di}}P_{Di} + \frac{\partial H_h^A}{\partial^A Q_{Di}}Q_{Di} + \frac{\partial H_h^A}{\partial^A Q_{Gi}^+}Q_{Gi}^+ + \frac{\partial H_h^A}{\partial^A Q_{Gi}^-}Q_{Gi}^- + \frac{\partial H_h^A}{\partial^A Q_{Di}}Q_{Di}\right)$$

$$+ \sum_{i\in PXS}\frac{\partial H_h^A}{\partial^A V_i}V_i + \sum_{\ell\in ND}\left(\frac{\partial H_h^A}{\partial^A x_\ell^+}x_\ell^+ + \frac{\partial H_h^A}{\partial^A x_\ell^-}x_\ell^-\right)$$

$$= \sum_{i\in PXS}\frac{\partial H_h^A}{\partial^A P_{Di}}P_{Di}^* + \sum_{i\in PXS}\frac{\partial H_h^A}{\partial^A Q_i}Q_{Di}^* + \sum_{i\in PXS}\frac{\partial H_h^A}{\partial^A V_i}V_i^* \qquad :\nu_h \qquad \forall h\in NN \quad (A.18)$$

Active and Reactive Generation and Voltage Bounds and Load Settings

$$-P_{Gi}^+ \ge -P_{Gi}^{max} + P_{Gi}^* \qquad\qquad :\upsilon_{Pi}^+ \qquad \forall i\in PXS \quad (A.19)$$

$$-P_{Gi}^- \ge P_{Gi}^{min} - P_{Gi}^* \qquad\qquad :\upsilon_{Pi}^- \qquad \forall i\in PXS \quad (A.20)$$

$$-Q_{Gi}^+ \ge -Q_{Gi}^{max} + Q_{Gi}^* \qquad\qquad :\upsilon_{Qn}^+ \qquad \forall i\in PXS \quad (A.21)$$

$$-Q_{Gi}^- \ge Q_{Gi}^{min} - Q_{Gi}^* \qquad\qquad :\upsilon_{Qn}^- \qquad \forall i\in PXS \quad (A.22)$$

$$-V_i \ge -V_i^{max} \qquad\qquad :\upsilon_{Vi}^+ \qquad \forall i\in PXS \quad (A.23)$$

$$V_i \ge V_i^{min} \qquad\qquad :\upsilon_{Vi}^- \qquad \forall i\in PXS \quad (A.24)$$

$$P_{Di} = P_{Di}^* \qquad\qquad :\beta_{Pi} \qquad \forall i\in PXS \quad (A.25)$$

$$Q_{Di} = Q_{Di}^* \qquad\qquad :\beta_{Qi} \qquad \forall i\in PXS \quad (A.26)$$

$$\underset{\substack{\nu^{NN}\, \nu_P^{+PXS}\, \nu_Q^{+PXS}\, \nu_V^{+PXS}\geq \mathbf{0} \\ \nu_P^{-PXS}\, \nu_Q^{-PXS}\, \nu_V^{-PXS}\geq \mathbf{0} \\ \lambda_P,\lambda_Q,\mu_Q^{PV}\mu_V^{PQ}\beta_P^{PXS}\beta_Q^{PXS}}}{\textbf{Maximise}} \quad \lambda_P\left(L_P^* + \sum_{i\in PX}\frac{\partial L_P^A}{\partial^A P_i}P_{Di}^* + \sum_{i\in PQ}\frac{\partial L_P^A}{\partial^A Q_i}Q_{Di}^* - \sum_{i\in PVS}\frac{\partial L_P^A}{\partial^A V_i}V_i^* - \sum_{i\in PXS}P_{Gi}^* \right)$$

$$+ \lambda_Q\left(L_Q^* + \sum_{i\in PX}\frac{\partial L_Q^A}{\partial^A P_i}P_{Di}^* + \sum_{i\in PQ}\frac{\partial L_Q^A}{\partial^A Q_i}Q_{Di}^* - \sum_{i\in PVS}\frac{\partial L_Q^A}{\partial^A V_i}V_i^* - \sum_{i\in PXS}Q_{Gi}^* \right)$$

$$+ \sum_{n\in PV}\mu_{Qn}\left(-Q_{Dn}^* + \sum_{i\in PX}\frac{\partial Q_n^A}{\partial^A P_i}P_{Di}^* + \sum_{i\in PQ}\frac{\partial Q_n^A}{\partial^A Q_i}Q_{Di}^* - \sum_{i\in PVS}\frac{\partial Q_n^A}{\partial^A V_i}V_i^* \right)$$

$$+ \sum_{n\in PQ}\mu_{Vn}\left(V_n^* + \sum_{i\in PX}\frac{\partial V_n^A}{\partial^A P_i}P_{Di}^* + \sum_{i\in PQ}\frac{\partial V_n^A}{\partial^A Q_i}Q_{Di}^* - \sum_{i\in PVS}\frac{\partial V_n^A}{\partial^A V_i}V_i^* \right)$$

$$+ \sum_{h\in NN}\nu_h\left(\sum_{i\in PXS}\frac{\partial H_h^A}{\partial^A P_{Di}}P_{Di}^* + \sum_{i\in PXS}\frac{\partial H_h^A}{\partial^A Q_i}Q_{Di}^* + \sum_{i\in PXS}\frac{\partial H_h^A}{\partial^A V_i}V_i^* \right)$$

$$+ \sum_{i\in PXS}\left(\upsilon_{Pi}^+\left(-P_{Gi}^{max}+P_{Gi}^*\right)+\upsilon_{Pi}^-\left(P_{Gi}^{min}-P_{Gi}^*\right)\right) + \sum_{i\in PXS}\left(\upsilon_{Qi}^+\left(-Q_{Gi}^{max}+Q_{Gi}^*\right)+\upsilon_{Qi}^-\left(Q_{Gi}^{min}-Q_{Gi}^*\right)\right)$$

$$+ \sum_{i\in PXS}\left(-\upsilon_{Vi}^+V_i^{max}+\upsilon_{Vi}^-V_i^{min}\right) + \sum_{i\in PXS}\beta_{Pi}P_{Di}^* + \sum_{i\in PXS}\beta_{Qi}Q_{Di}^*$$

(A.27)

Constraint set begins on following page

Subject to:

Nodal Prices Defined by OPF Demand Terms

$$-\lambda_P\left(1-\frac{\partial L_P^A}{\partial^A P_i}\right)+\lambda_Q\frac{\partial L_Q^A}{\partial^A P_i}+\sum_{n\in PV}\left(\mu_{Qn}\frac{\partial Q_n^A}{\partial^A P_i}\right)+\sum_{n\in PQ}\left(\mu_{Vn}\frac{\partial V_n^A}{\partial^A P_i}\right)+\sum_{h\in NN}\left(\nu_h\frac{\partial H_h^A}{\partial^A P_{Di}}\right)+\beta_{Pi}=0 \qquad :P_{Di} \qquad \forall i\in PX \quad \text{(A.28)}$$

$$-\lambda_P+\sum_{h\in NN}\left(\nu_h\frac{\partial H_h^A}{\partial^A P_{Ds}}\right)+\beta_{Ps}=0 \qquad :P_{Ds} \qquad\qquad \text{(A.29)}$$

$$\lambda_P\frac{\partial L_P^A}{\partial^A Q_i}-\lambda_Q\left(1-\frac{\partial L_Q^A}{\partial^A Q_i}\right)+\sum_{n\in PV}\left(\mu_{Qn}\frac{\partial Q_n^A}{\partial^A Q_i}\right)+\sum_{n\in PQ}\left(\mu_{Vn}\frac{\partial V_n^A}{\partial^A Q_i}\right)+\sum_{h\in NN}\left(\nu_h\frac{\partial H_h^A}{\partial^A Q_{Di}}\right)+\beta_{Qi}=0 \qquad :Q_{Di} \qquad \forall i\in PQ \quad \text{(A.30)}$$

$$-\lambda_Q-\mu_{Qi}+\sum_{h\in NN}\left(\nu_h\frac{\partial H_h^A}{\partial^A Q_{Di}}\right)+\beta_{Qi}=0 \qquad :Q_{Di} \qquad \forall i\in PV \quad \text{(A.31)}$$

$$-\lambda_Q+\sum_{h\in NN}\left(\nu_h\frac{\partial H_h^A}{\partial^A Q_{Ds}}\right)+\beta_{Qs}=0 \qquad :Q_{Ds} \qquad\qquad \text{(A.32)}$$

<u>Constraint set continues on following page</u>

Floor And Ceiling Constraints Set by Generating Unit Costs

$$\pm \lambda_P \left(1 - \frac{\partial L_P^A}{\partial^A P_i} \right) \mp \lambda_Q \frac{\partial L_Q^A}{\partial^A P_i} \mp \sum_{n \in PV} \left(\mu_{Qn} \frac{\partial Q_n^A}{\partial^A P_i} \right) \mp \sum_{n \in PQ} \left(\mu_{Vn} \frac{\partial V_n^A}{\partial^A P_i} \right) + \sum_{h \in NN} \left(v_h \frac{\partial H_h^A}{\partial^A P_{Gi}^\pm} \right) - v_{Pi}^\pm \leq \pm c_{Pi}^\pm \quad : P_{Gi}^\pm \qquad \forall i \in PX \quad \text{(A.33)}$$

$$\pm \lambda_P + \sum_{h \in NN} \left(v_h \frac{\partial H_h^A}{\partial^A P_{Ps}^\pm} \right) - v_{Ps}^\pm \leq \pm c_{Ps}^\pm \quad : P_{Gs}^\pm \qquad \text{(A.34)}$$

$$\mp \lambda_P \frac{\partial L_P^A}{\partial^A Q_i} \pm \lambda_Q \left(1 - \frac{\partial L_Q^A}{\partial^A Q_i} \right) \mp \sum_{n \in PV} \left(\mu_{Qn} \frac{\partial Q_n^A}{\partial^A Q_i} \right) \mp \sum_{n \in PQ} \left(\mu_{Vn} \frac{\partial V_n^A}{\partial^A Q_i} \right) + \sum_{h \in NN} \left(v_h \frac{\partial H_h^A}{\partial^A Q_{Gi}^\pm} \right) - v_{Qi}^\pm \leq \pm c_{Qi}^\pm \quad : Q_{Gi}^\pm \qquad \forall i \in PQ \quad \text{(A.35)}$$

$$\pm \lambda_Q \pm \mu_{Qi} + \sum_{h \in NN} \left(v_h \frac{\partial H_h^A}{\partial^A Q_{Gi}^\pm} \right) - v_{Qi}^\pm \leq \pm c_{Qi}^\pm \quad : Q_{Gi}^\pm \qquad \forall i \in PV \quad \text{(A.36)}$$

$$\pm \lambda_Q + \sum_{h \in NN} \left(v_h \frac{\partial H_h^A}{\partial^A Q_{Gs}^\pm} \right) - v_{Qs}^\pm \leq \pm c_{Qs}^\pm \quad : Q_{Gs}^\pm \qquad \text{(A.37)}$$

Voltage Constraint Cost Relationships

$$- \lambda_P \frac{\partial L_P^A}{\partial^A V_i} - \lambda_Q \frac{\partial L_Q^A}{\partial^A V_i} - \sum_{n \in PV} \left(\mu_{Qn} \frac{\partial Q_n^A}{\partial^A V_i} \right) - \sum_{n \in PQ} \left(\mu_{Vn} \frac{\partial V_n^A}{\partial^A V_i} \right) + \sum_{h \in NN} \left(v_h \frac{\partial H_h^A}{\partial^A V_i} \right) - v_{Vi}^+ + v_{Vi}^- = 0 \quad : V_i \qquad \forall i \in PVS \quad \text{(A.38)}$$

$$\mu_{Vi} + \sum_{h \in NN} \left(v_h \frac{\partial H_h^A}{\partial^A V_i} \right) - v_{Vi}^+ + v_{Vi}^- = 0 \quad : V_i \qquad \forall i \in PQ \quad \text{(A.39)}$$

General Variable Constraints Cost Relationships

$$\sum_{h \in NN} \left(v_h \frac{\partial H_h^A}{\partial^A x_\ell^\pm} \right) \leq \pm c_{x\ell}^\pm \quad : x_\ell \qquad \forall \ell \in ND \quad \text{(A.40)}$$

A.3. The Dual of the Linearised OPF

Standard mathematical programming theory[181] can be used to form the dual of equations (A.13) to (A.26). This dual is described by equations (A.27) to (A.40). The primal variables corresponding to each of the dual constraints are indicated. Note that for compactness that the equations of (A.33) to (A.37) each simultaneous represent the dual constraints of both a marginal increase and decrease in generation.

A.3.1 Theoretical Properties of the Dual OPF

The following theorems summarise some important properties of the dual OPF.

Theorem A.1: The dual of the linearised OPF is the dual of the non-linear OPF if the non-linear OPF is convex and its solution is optimal.

Theorem A.2: If the solution to the OPF is given then any set of dual prices will suffice as long as they satisfy the dual constraints and the complimentary slackness conditions. That is, the dual objective function is arbitrary.

Proof of Theorems A.1 and A.2.

Theorems A.1 and A.2 can be proved using the Kuhn-Tucker conditions but the following proof, which employs the Wolfe Dual (Wolfe, 1961) is somewhat more explicit. We use the form of the Wolfe Dual described by Balinsky and Baumol (1968).

Consider the non-linear primal problem:

$$\underset{x}{\textbf{Minimise}} \quad f(x) \equiv f(x_1, \ldots x_n)$$

subject to:

$$g_1(x) \quad \equiv \quad g_1(x_1, \ldots, x_n) \quad \geq \quad c_1 \qquad \qquad : v_1$$
$$\vdots \qquad\qquad \vdots \qquad\qquad\quad \vdots \qquad\qquad \vdots$$
$$g_m(x) \quad \equiv \quad g_m(x_1, \ldots, x_n) \quad \geq \quad c_m \qquad \qquad : v_m$$
$$x_1 \geq 0, \ldots, x_n \geq 0$$

The Wolfe Dual of this is a non-linear dual, formed without needing to explicitly linearised the primal problem. This dual has the form:

---

[181]   See Bazaraa and Jarvis (1977), for example.

$$\textbf{Maximise}_{v} \quad f(x) + \sum_{j=1}^{m} v_j \left(c_j - g_j(x) \right) - \sum_{i=1}^{n} x_i \left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} v_j \frac{\partial g_j}{\partial x_i} \right)$$

subject to:

$$
\begin{array}{ccccccc}
v_1 \dfrac{\partial g_1}{\partial x_1} + & \cdots & + v_m \dfrac{\partial g_m}{\partial x_1} & \leq & \dfrac{\partial f}{\partial x_1} & & : x_1 \\
\vdots & & \vdots & & \vdots & & \vdots \\
v_1 \dfrac{\partial g_1}{\partial x_n} + & \cdots & + v_m \dfrac{\partial g_m}{\partial x_n} & \leq & \dfrac{\partial f}{\partial x_n} & & : x_m
\end{array}
$$

$$v_1 \geq 0, \ldots, v_n \geq 0$$

At the optimal OPF solution the constraints of this dual problem are identical to the constraints which would be produced had we formed the dual of the linearised primal. Furthermore, given the primal solution is already known, the objective function reduces to:

$$\textbf{Maximise}_{v} \quad f(x^*) + \sum_{j=1}^{m} v_j \left(c_j - g_j(x^*) \right) - \sum_{i=1}^{n} x_j^* \left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} v_j \frac{\partial g_j}{\partial x_i} \right)$$

But $f(x^*)$ is a constant and, since the remaining terms are just the complementary slackness conditions, we are left with:

$$\sum_{j=1}^{m} v_j \left(c_j - g_j(x^*) \right) - \sum_{i=1}^{n} x_j^* \left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} v_j \frac{\partial g_j}{\partial x_i} \right) = 0$$

That is, any prices v which satisfy both the dual constraints and this complementary slackness requirement are feasible prices. Consequently, and as long as the complementary slackness conditions are enforced, the dual objective function is redundant, and we are justified in introducing any new pricing objective function. Hence we can re-express the dual as[182]:

$$\textbf{Minimise}_{v} \quad \text{Arbitrary Objective Function}$$

subject to:

$$
\begin{array}{ccccc}
\langle v_1 \rangle \left\langle \dfrac{\partial g_1}{\partial x_1} \right\rangle + & \cdots & + \langle v_m \rangle \left\langle \dfrac{\partial g_m}{\partial x_1} \right\rangle & \leq & \dfrac{\partial f}{\partial x_1} \\
\vdots & & \vdots & & \vdots \\
\langle v_1 \rangle \left\langle \dfrac{\partial g_1}{\partial x_n} \right\rangle + & \cdots & + \langle v_m \rangle \left\langle \dfrac{\partial g_m}{\partial x_n} \right\rangle & \leq & \dfrac{\partial f}{\partial x_n}
\end{array}
$$

$$v_1 \geq 0, \ldots, v_n \geq 0$$

---

[182]   The choice of minimising the objective function is as arbitrary as the objective function itself.

Here $\langle z \rangle$ indicates that the term z is zero unless the primal constraint it corresponds to is binding[183]. Provided that primal constraints which are binding can readily be identified it should be straight-forward to remove these terms from the pricing problem. While there will generally be only one solution to this problem, non-differentiable points in the OPF can result in some freedom with which values can be assigned to the dual values. As a result multiple dual solutions may exist for a given primal solution (as multiple primal solutions may exist for a given dual solution). This is where the use of a new pricing objective function is useful, as the original dual objective function cannot distinguish between degenerate solutions. We have not suggested an explicit objective function as this is ultimately a policy issue. In his formulation, Hogan suggests an objective function which minimised the transmission rents collected by the network, while Read and Ring (1995b) suggest maximising the transmission rents as a non-distortionary means for recovering some of the network fixed costs.

As the dual constraints in the non-linear dual are identical to the dual constraints for a linearised primal problem, and since the dual objective function is arbitrary, we have proved both Theorems A.1 and A.2.

A.3.2 Simplifying the Dual OPF.

A number of simplifications can be made to the dual OPF. These are:

- The general constraints involve the only functions which are dependent on the swing bus active and reactive power generation and load levels. As a consequence, equation (A.29) can be represented in the same form as equation (A.28), but where all the derivatives with respect to active power demand, other than for the general constraint, are zero. Similarly, equation (A.32) can be expressed in the same form as (A.30). Hence (A.28), (A.29), (A.30), and (A.32) can be written as:

$$\beta_{Pi} = \lambda_P \left(1 - \frac{\partial L_P^A}{\partial^A P_i} \right) - \lambda_Q \frac{\partial L_Q^A}{\partial^A P_i} - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A P_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A P_i} - \sum_{h \in NN} \nu_h \frac{\partial H_h^A}{\partial^A P_{Di}}$$

$$\forall i \in PXS \quad (A.41)$$

$$\beta_{Qi} = -\lambda_P \frac{\partial L_P^A}{\partial^A Q_i} + \lambda_Q \left(1 - \frac{\partial L_Q^A}{\partial^A Q_i}\right) - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A Q_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A Q_i} - \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A Q_{Di}}$$

$$\forall i \in PQS \quad (A.42)$$

- Equation (A.31) can be re-arranged as:

$$\mu_{Qi} = \beta_{Qi} - \lambda_Q + \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A Q_{Di}} \qquad \forall i \in PV \quad (A.43)$$

- We can define a price for voltage, in terms of the value of voltage capacity[184], as:

$$\beta_{Vi} = v_{Vi}^+ - v_{Vi}^- \qquad i \in PXS \quad (A.44)$$

And hence re-express (A.38) and (A.39) as:

$$\beta_{Vi} = -\lambda_P \frac{\partial L_P^A}{\partial^A V_i} - \lambda_Q \frac{\partial L_Q^A}{\partial^A V_i} - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A V_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A V_i} - \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A V_i}$$

$$\forall i \in PVS \quad (A.45)$$

$$\mu_{Vi} = \beta_{Vi} - \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A V_i} \qquad \forall i \in PQ \quad (A.46)$$

- By substituting the equations of (A.41) into (A.33) and (A.34) and manipulating the results we get:

$$c_{Pi}^- - v_{Pi}^- + \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A P_{Gi}^-} \leq \beta_{Pi} + \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A P_{Di}} \leq c_{Pi}^+ + v_{Pi}^+ - \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A P_{Gi}^+}$$

$$\forall i \in PXS \quad (A.47)$$

Similarly, substituting equations from (A.42) into (A.35) and (A.37), and from (A.43) into (A.36), and manipulating the results we get:

$$c_{Qi}^- - v_{Qi}^- + \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A Q_{Gi}^-} \leq \beta_{Qi} + \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A Q_{Di}} \leq c_{Qi}^+ + v_{Qi}^+ - \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A Q_{Gi}^+}$$

$$\forall i \in PXS \quad (A.48)$$

- The two equation sets represented by (A.40) can be manipulated to give:

$$c_{x\ell}^- + \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A x_\ell^-} \leq 0 \leq c_{x\ell}^+ - \sum_{h \in NN} v_h \frac{\partial H_h^A}{\partial^A x_\ell^+} \qquad \forall \ell \in ND \quad (A.49)$$

---

[184] This is discussed further in Chapter 4.

- Finally, using Theorem A.2, we introduce an arbitrary pricing objective function and introduced the $\langle z \rangle$ notation described above for those terms which correspond to inequality constraints which may not be binding at optimality.

The simplified form of the dual OPF is described by equations (A.50) to (A.59) as well as by (4.15) to (4.24).

$$\underset{\substack{\boldsymbol{v}^{NN}, \boldsymbol{v}_P^{+PXS}, \boldsymbol{v}_Q^{+PXS}, \boldsymbol{v}_V^{+PXS} \geq \mathbf{0} \\ \boldsymbol{v}_P^{-PXS}, \boldsymbol{v}_Q^{-PXS}, \boldsymbol{v}_V^{-PXS} \geq \mathbf{0} \\ \lambda_P, \lambda_Q, \boldsymbol{\mu}_Q^{PV}, \boldsymbol{\mu}_V^{PQ}, \boldsymbol{\beta}_P^{PXS}, \boldsymbol{\beta}_Q^{PXS}}}{\textbf{Minimise}} \qquad \text{Arbitrary Objective Function} \tag{A.50}$$

subject to:

$$\beta_{Pi} = \lambda_P\left(1 - \frac{\partial L_P^A}{\partial^A P_i}\right) - \lambda_Q \frac{\partial L_Q^A}{\partial^A P_i} - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A P_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A P_i} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Di}} \right\rangle \qquad \forall i \in PXS \quad (A.51)$$

$$\beta_{Qi} = -\lambda_P \frac{\partial L_P^A}{\partial^A Q_i} + \lambda_Q\left(1 - \frac{\partial L_Q^A}{\partial^A Q_i}\right) - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A Q_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A Q_i} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Di}} \right\rangle \qquad \forall i \in PQS \quad (A.52)$$

$$\beta_{Vi} = -\lambda_P \frac{\partial L_P^A}{\partial^A V_i} - \lambda_Q \frac{\partial L_Q^A}{\partial^A V_i} - \sum_{n \in PV} \mu_{Qn} \frac{\partial Q_n^A}{\partial^A V_i} - \sum_{n \in PQ} \mu_{Vn} \frac{\partial V_n^A}{\partial^A V_i} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A V_i} \right\rangle \qquad \forall i \in PVS \quad (A.53)$$

$$\mu_{Qi} = \beta_{Qi} - \lambda_Q + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Di}} \right\rangle \qquad \forall i \in PV \quad (A.54)$$

$$\mu_{Vi} = \beta_{Vi} - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A V_i} \right\rangle \qquad \forall i \in PQ \quad (A.55)$$

$$c_{Pi}^- - \langle v_{Pi}^- \rangle + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Gi}^-} \right\rangle \leq \beta_{Pi} + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Di}} \right\rangle \leq c_{Pi}^+ + \langle v_{Pi}^+ \rangle - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A P_{Gi}^+} \right\rangle \qquad \forall i \in PXS \quad (A.56)$$

$$c_{Qi}^- - \langle v_{Qi}^- \rangle + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Gi}^-} \right\rangle \leq \beta_{Qi} + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Di}} \right\rangle \leq c_{Qi}^+ + \langle v_{Qi}^+ \rangle - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A Q_{Gi}^+} \right\rangle \qquad \forall i \in PXS \quad (A.57)$$

$$\beta_{Vi} = \langle v_{Vi}^+ \rangle - \langle v_{Vi}^- \rangle \qquad \forall i \in PXS \quad (A.58)$$

$$c_{x\ell}^- + \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A x_\ell^-} \right\rangle \leq 0 \leq c_{x\ell}^+ - \sum_{h \in NN} \langle v_h \rangle \left\langle \frac{\partial H_h^A}{\partial^A x_\ell^+} \right\rangle \qquad \forall \ell \in ND \quad (A.59)$$

Appendix B

Notation and Notational Conventions

B.1. Introduction

This appendix summarises the notation and terminology used in this thesis.

B.2. General Terminology

bus or bus-bar   A collection of nodes used in an engineering representation of a power system.

consumers   Those market participants who consume power.

dispatcher   The entity which coordinates the dispatch of the power system.   The dispatcher may also be the grid owner.

ex post   After the event.

ex ante   Prior to the event.

generating company   The company that operates one or more generating stations.

grid owner   The owner of a high voltage transmission grid.

generating station   A set of one or more generating units.

generating unit   A physical generating turbine.
load   The total demand of the consumers at a node or overall.

node A point in the network to which a load, generating unit or transformer is attached. One device may be attached at several nodes, these nodes collectively forming a bus or bus-bar.   We often assume that there is only one generating unit at a node, or, when there are multiple units, that all the units there behave identically.

B.3.   Abbreviations and Units

$/MW   Dollars per megawatt[185].

AC   Alternating current.

AGC Automatic generation controls

c/kW Cents per kilowatt[186].

CCOPF   Contingency constrained optimal power flow.

CP   Compensation payment.

DC   Direct current.

EAP Effective actual profits earned by a generating unit. EAP=PWC+CP.

HVDC   High voltage direct current.

Hz   Hertz.

ICR   Industry capacity right.

kV   Kilovolt.

LRMC   Long run marginal cost.

MW Megawatt.

MOM   Maximum outage magnitude (for ODC).

MRO formulation   A multiple period best compromise pricing formulation including running overheads.

---

[185]   Rather than describing prices as being for power usage over some defined period, such as an hour, we assume an arbitrary, unitless, interval. Hence instead of stating that $5/MWh was paid for 3MWh, the total payment being $15, we simply state that $5/MW was paid for 3MW, the total payment being $15.

[186]   Rather than describing prices as being for power usage over some defined period, such as an hour, we assume an arbitrary, unitless, interval. Hence instead of stating that $5/MWh was paid for 3MWh, the total payment being $15, we simply state that $5/MW was paid for 3MW, the total payment being $15.

MSRO formulation   A multiple period best compromise pricing formulation including start up costs and running overheads.

MVA Mega-Volt-Amperes.

ODC Outage duration curve.

OOMOD   Out of merit order dispatch.

OPF Optimal power flow.

ORT model   Optimal Reserve Targeting model.

PDC Price duration curve.

RO formulation A single period best compromise pricing formulation including running overheads.

SRMC   Short run marginal cost.

TCOT   Total cumulative outage time (for ODC).

B.4.   Mathematical Notation

Notation is listed according to the chapter in which it is first used.   Notational conventions used by Hogan (1991) are also presented where relevant.

Chapter 4

Attribute	This Document	Hogan
Set of all nodes	*PXS*	
Set of non-swing bus nodes	*PX*	
Set of all voltage controlled nodes	*PVS*	
Set of non-swing bus voltage controlled nodes	*PV*	
Set of reactive power controlled nodes	*PQ*	
Set of general OPF constraints	*NN*	

Chapter 4 (continued)

Attribute	This Document	Hogan
Set of general OPF dependent variables which are not otherwise specified	ND	
Indicator of PVQ basis representation	A	
Indicator of PQ basis representation	B	
The "natural" electrical basis involving only voltage magnitudes and phase angles.	$V\theta$	
The union of sets PQ and the swing bus.	PQS	
Node index	i,j,n	i,j
The swing bus index	$i=s$ or "set" S	s
Index for general OPF variables	ℓ	
Index for general OPF constraints	h	
Active power generation at node i	P_{Gi}	gp
Reactive power generation at node i	Q_{Gi}	gQ
Active power generation at the swing bus	P_{Gs}	gps
Reactive power generation at the swing bus	Q_{Gs}	gQs
Reactive power demand at node i	Q_{Di}	dQi, dQ
Active power demand at node i	P_{Di}	dpi
Levels of active power demand at node i set exogenously to OPF	P_{Di}^{set}	
Levels of reactive power demand at node i set exogenously to OPF	Q_{Di}^{set}	
Active power net injection at node i	P_i	ypi
Reactive power net injection at node i	Q_i	yQi
Voltage magnitude at node i	V_i	V
Phase angle at node i	θ_i	δ

Chapter 4 (continued)

Attribute	This Document	Hogan
Vector of nodal voltage levels	$\boldsymbol{V}$	
Vector of nodal active power demand levels	$\boldsymbol{P}_D$	dp
Vector of nodal reactive power demand levels	$\boldsymbol{Q}_D$	
Vector of nodal active power generation levels	$\boldsymbol{P}_G$	
Vector of nodal active power generation levels	$\boldsymbol{Q}_G$	
The vector of general variables x_ℓ	$\boldsymbol{x}$	
A shadow price relating to general variable x_ℓ	$\alpha_{x\ell}$	
Total active power losses	L_P	L, L_P, L_p
Total reactive power losses	L_Q	L_Q, L_q
General OPF constraint h	H_h	
General OPF variable ℓ	x_ℓ	
Generation cost function	$C(.)$	
Cost of last unit of active power generation at node i	c_{Pi}^-	c_i
Cost of next unit of active power generation at node i	c_{Pi}^+	w_i
Cost of last unit of reactive power generation at node i	c_{Qi}^-	
Cost of next unit of reactive power generation at node i	c_{Qi}^+	
Cost of last unit of general variable x_ℓ	$c_{x\ell}^-$	
Cost of next unit of general variable x_ℓ	$c_{x\ell}^+$	
Marginal value of active power	λ_P	σ_P (opposite sign convention)
Marginal value of reactive power	λ_Q	σ_Q (opposite sign convention)

Chapter 4 (continued)

Attribute	This Document	Hogan
Active power price at node i	β_{Pi}	P_{Pi}
Reactive power price at node i	β_{Qi}	P_{Qi}
Voltage price at node i	β_{Vi}	
Swing bus active power price	β_{Ps}	P_{Ps}
Swing bus reactive power price	β_{Qs}	P_{Qs}
Voltage price at swing bus	β_{Vs}	
Shadow price on constraint defining dependent reactive power injections at node i	μ_{Qi}	
Shadow price on constraint defining dependent voltage at node i	μ_{Vi}	θ_V
Upper and lower bound shadow prices for active power generation at node i	$\upsilon_{Pi}^+, \upsilon_{Pi}^-$	
Upper and lower bound shadow prices for reactive power generation at node i	$\upsilon_{Qi}^+, \upsilon_{Qi}^-$	
Upper and lower bound shadow prices for voltage at node i	$\upsilon_{Vi}^+, \upsilon_{Vi}^-$	τ_1, τ_2
Marginal (active power) profit of a generating unit.	$\upsilon_{Pi}^+ - \upsilon_{Pi}^-$	ϕ
Shadow price on general constraint h	ν_h	
Only non-zero if the primal constraint corresponding to the term z is binding	$< z >$	$< z >$
Indicates the lower bound of a variable.	*min* (superscripted)	
Indicates the upper bound of a variable.	*max* (superscripted)	
Observed value, symbol	*(superscript)	* (superscript)
For all	$\forall$	
Partial derivative with respect to, for example, representation A	$\dfrac{\partial}{\partial^A}$	$\dfrac{\partial}{\partial}, \nabla$

Chapter 5

Attribute	This Document	Hogan
Set of transmission lines	K	
Transmission lines index	k (absolute circuit index)	k (kth circuit between i and j)
Conventional "sending" end of transmission line	$i(k)$	
Conventional "receiving" end of transmission line	$j(k)$	
Average active power flow on line k	$\overline{P}_k$	zPk
average reactive power flow on line k	$\overline{Q}_k$	zQk
Maximum and minimum active power flows on line k	$\overline{P}_k^{max}, \overline{P}_k^{min}$	
Maximum and minimum reactive power flows on line k	$\overline{Q}_k^{max}, \overline{Q}_k^{min}$	
Thermal transmission constraint right-hand side	T_k^{max}	
Shadow price on maximum active power megawatt flow on line k	$\eta_{\overline{P}k}$	θP
Shadow price on maximum reactive power megawatt flow on line k	$\eta_{\overline{Q}k}$	θQ
Shadow price on thermal transmission constraint on line k	χ_k	μ
Upper and lower bound shadow prices for active power flows on line k	$\upsilon_{\overline{P}k}^+, \upsilon_{\overline{P}k}^-$	
Upper and lower bound shadow prices for reactive power flows on line k	$\upsilon_{\overline{Q}k}^+, \upsilon_{\overline{Q}k}^-$	
Marginal node	m	
The resistance between the origin and point X on a loop.	R_X	
The resistance between nodes i and j.	R_{ij}	
The total resistance around a loop.	R_T	

Chapter 6

Attribute	This Document	Hogan
The set of reserve constraints	RC	
Reserve constraint h	$R_h(.)$	
Derivative of reserve constraint h with respect to increasing active power generation at node i	ρ_{hi}^+	
Derivative of reserve constraint h with respect to decreasing active power generation at node i	ρ_{hi}^-	
Contribution of generating unit i to reserve constraint h.	R_{hi}	
Component of contribution of generating unit i to reserve constraint h which is independent of the output of generating unit i.	a_{hi}	
Component of reserve constraint h which is independent of the output of any generating unit.	R_{h0}	
Shadow price on reserve constraint h (no h if only one constraint)	γ_h	
Transmission price between two nodes	TP	TP
Total profit of generating unit at node i.	Π_i	

Chapter 7

Attribute	This Document	Hogan
Set of all primal variables	NDI	
Set of all primal constraints	CS	
Amount of resource h used by variable x_ℓ.	$a_{h\ell}$	
Amount of resource h available.	b_h	
A change in value of general variable ℓ	Δx_ℓ	
Shadow price on general primal constraint h.	ς_h	

Chapter 7 (continued)

Attribute	This Document	Hogan
Value of general primal variable ℓ.	$\beta_{x\ell}$	
Objective function optimal value for the first phase of best compromise pricing.	Ω	
Total capacity rights sold for general variable x_ℓ.	$x_\ell^{\text{capacity right}}$	
Capacity of general variable x_ℓ assumed by the dispatcher.	CAP_ℓ	

Chapter 8

Attribute	This Document	Hogan
Set of time periods.	T	
Time period	t (superscript)	
Shadow price on inter-temporal ramping constraint for generating unit i.	τ_{Pi}	
Shadow price on inter-temporal energy constraint for generating unit i.	ξ_{Pi}	

Chapter 9

Attribute	This Document	Hogan
The set of contingency states	U	
The set of generators.	PG	
Index for stochastic contingency states.	u(superscript)	
Indicator of the base operational state.	0(superscript)	
Probability of base state occurring.	π^0	
Probability of contingency state u occurring.	π^u	

Chapter 9 (continued)

Attribute	This Document	Hogan
A 0-1 parameter which takes a value of 1 if generating unit i is operational in state u	δ_i^u	
Spinning reserve from generating unit i utilised in contingency state u	p_{Ri}^u	
Load curtailment utilised in contingency state u.	p_C^u ·	
Blackout utilised in contingency state u.	p_B^u	
Marginal cost of calling on load curtailment.	c_C	
Marginal cost of blackout (system shortage cost).	c_B	
Spinning reserve capacity from generating unit i scheduled.	P_{Ri}	
Load curtailment capacity scheduled.	P_C	
Spinning reserve capacity from generating unit i made available to the dispatcher.	P_{Ri}^{max}	
Load curtailment capacity made available to the dispatcher.	P_C^{max}	
The marginal value of base state generation.	λ_P^0	
The marginal value of reserve provided in contingency state u.	λ_P^u	
Value of the extra energy that can be called upon from generating unit i in contingency state u.	ω_{Ri}^u	
Value of the load curtailment that can called open in contingency state u.	ω_C^u	
Expected value of load curtailment capacity.	υ_C^+	
Expected value of spinning reserve capacity.	υ_{Ri}^+	
The minimum λ_P^u value.	λ_P^{min}	

Chapter 9 (continued)

Attribute	This Document	Hogan
The fuel cost of the generating unit which sets λ_P^{min} (i.e. $c_{PR} = \lambda_P^{min}$)	c_{PR}	
The fuel cost of the marginal generating unit.	c_{PM}	

Chapter 10

Attribute	This Document	Hogan
The per period running overhead of generating unit i.	f_i	
A 0-1 parameter which equals 1 if generating unit i is committed during a period.	δ_{fi}	
Minimum average cost of generating unit i.	$\hat{c}_{Pi}$	
Observed actual average cost of generating unit i.	c_{Pi}^o	
A compensation term which accounts for the difference between the minimum and actual average cost of generating unit i.	Ω_{fi}	
The start up cost of generating unit i.	F_i	
A 0-1 parameter which equals 1 if generating unit i is committed during the day.	δ_{Fi}	
A compensation term to recover the start up costs of generating unit i not recovered via the spot prices.	U_i	

Appendix One

Attribute	This Document	Hogan
Incremental increase in active power generation at node i.	$P_{Gi}^+ \geq 0$	$\Delta g_{Pi} \geq 0$
Incremental decrease in active power generation at node i.	$P_{Gi}^- \geq 0$	$g_{Pi} \leq g_{Pi}^*$
Incremental increase in reactive power generation at node i.	$Q_{Gi}^+ \geq 0$	
Incremental decrease in reactive power generation at node i.	$Q_{Gi}^- \geq 0$	
Incremental increase in general OPF variable ℓ	$x_{\ell}^+ \geq 0$	
Incremental increase in general OPF variable ℓ	$x_{\ell}^- \geq 0$	