Simulation of Electricity Markets using Agent-Based Computational Learning

Fernando Manuel Soares Mota Siciliani de Oliveira

Thesis submitted to the University of London
for the degree of Doctor of Philosophy

LONDON BUSINESS SCHOOL

SEPTEMBER 2003

ABSTRACT

The purpose of this research is to conduct an analysis of how agent-based computational learning may contribute to a better understanding of pricing policies and strategic management of plant portfolio in electricity markets. The contributions of this thesis are methodological and theoretical with applications in policy analysis for electricity markets.

At a policy level, this thesis applies agent-based simulation to the analysis of the impact of market design on the players' strategies and on the industry's performance as a whole. This represents the first detailed study of the New Electricity Trading Arrangements (NETA) in the England and Wales (E&W) electricity market, giving insights into the implications of NETA before its introduction. Further, this thesis addresses the issue of dominant position abuse by individual players in electricity markets. The context is a real application to the E&W electricity market as part of a Competition Commission Inquiry. The research contributions are in the areas of both market power and market design policy issues.

At a methodological level, this thesis presents two contributions: the Finite Automata Dynamic Game (FADG) and the Plant Trading Game. The FADG models learning and adaptation in *N*-player extensive form games of incomplete information, where

co-evolutionary automata learn and adapt together. The plant trading game is a large coordination game, simulating how players optimally learn and adapt in order to trade electricity plants.

At a theoretical level, this thesis presents three contributions. First, it develops a stylised model for conduct-evaluation in electricity markets, which is applied to the analysis of market power abuse and regulatory policy. Second, it studies plant trading within the context of a Cournot game. Third, it shows that, in an FADG, best response is a necessary but not a sufficient condition for rational behaviour.

ACKNOWLEDGEMENTS

I am forever indebted to my thesis supervisor, Prof. Derek Bunn, for his attentive and enlightened guidance that was so invaluable throughout this project.

Further, I would like to thank the scientific advice of several other people: Prof. Zeger DeGraeve, Prof. Victor DeMiguel, Prof. Scott Rockart and Prof. Andrew Scott.

At an institutional level, I would like to gratefully acknowledge the financial support by Fundação para a Ciência e a Tecnologia and Universidade do Porto (Portugal). I am also thankful to Prof. Manuel Costa for his friendly advice and for managing my relationship with Universidade do Porto.

Finally, at a personal level, I thank my family and friends for having made this journey much more enjoyable. Special thanks are due to my colleagues Martin, Maria, Jens, Raf, Richard and Rui, to my parents Aníbal and Dores and my siblings Rui and Lurdes. Finally, I thank Daphne for her enthusiasm and understanding, for her smile and the happiness she brought me, and for making my life beautiful during the course of my studies. To her I dedicate this thesis.

Fernando de Oliveira,
September 2003

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In several countries, the electricity industry is currently the object of a process of liberalisation and privatisation, which started back in the early nineties aiming at encouraging competition in the generation sector. The goal of the policy makers leading the different privatisation processes is to create a "reasonable" market by designing an industry where the costs of market imperfection are less than the costs of continuing regulation (Joskow, 1997).

Therefore, in general, liberalisation tends to replace the tight regulation of vertically integrated monopolies by the supervision of competitive markets (see Wilson, 2002, for a retrospective analysis of these processes). Further, the industry emerging from this process is characterised by decentralised decisions (regarding electricity pricing, plant maintenance planning and investment), increased uncertainty (the investment projects and pricing strategies are not public domain) and interaction between the different players (unbundled electricity companies and regulator).

If the goal of privatisation is to increase competition in the generation sector, the first experiences of these liberalised markets, however, have tended to emit contradictory signals regarding this evolution. If there is some evidence that the industry is now more efficient[1], there is also mounting evidence, in the new liberalised markets, that the few generation companies that exist tend to abuse their dominant positions, manipulating prices, withholding capacity and therefore reducing consumer welfare (e.g., Joskow and Khan, 2002; Wolfram, 1998).

A fundamental tool for analysing the liberalisation of electricity markets has been computer simulation, as it enables a beautiful mix of economic theory with the level of detail required by operational research and market design models (Roth, 2002). However, the different modelling techniques used so far contain several limitations that undermine their ability to be helpful in explaining or prescribing behaviour and (or) advising policy makers. System dynamic models (e.g., Larsen and Bunn, 1999) tend to capture well the overall behaviour of a system, but they have problems dealing with decentralised decision-making. On the other hand, agent-based models that assume perfect-rationality have problems capturing how players behave in the short-run, doing much better in explaining the long-term attractors of the system.

Next, section 1.1 presents some of the theoretical reasons for using agent-based simulation, and section 1.2 motivates the two main topics addressed in this thesis (firstly market design and market power, and secondly structural evolution) by looking at the England and Wales (E&W) electricity market.

[1] Salama (1997) shows how cultural change improved the efficiency of the management of nuclear plants in England and Wales.

1.1. Agent-Based Simulation

The main challenge faced by the scientific community working on bounded rationality is arguably the development of a model of rationality that can be used to understand human and organisational behaviour: this is also the task faced by agent-based simulation in economics and management science.

The main advantage of this approach is the development of models that can better mimic how humans behave, enabling the analysis of the adaptation process between equilibria. Empirical studies (e.g., Roth and Erev, 1995; Feltovich, 2000; Sarin and Vahid, 2001) show that models of bounded rationality and learning are more useful than the Nash equilibrium for predicting the behaviour of people, organisations or markets. At a theoretical level, Aumann (1997) gives several reasons for economists to look at models of bounded rationality. First, people do not optimise even in simple decision problems. Second, for many hard optimisation problems no solution is available within a useful time. Third, the analysis of models of rationality fails to explain human behaviour.

Thus, a first criticism of the perfect-rationality paradigm is that, in most situations, an agent does not know *a priori* the outcomes associated with his actions, he needs to learn them by interacting with the environment. A first solution is provided by using reinforcement learning, an agent may be able to learn the value of his actions by interacting with the environment (Bertsekas and Tsitsiklis, 1996). Unfortunately, these algorithms tend to require an extremely large number of interactions, in order for the agent to learn the best policy[2]; moreover, reinforcement learning assumes that an agent can try infinitely often any one of his possible actions.

A second criticism of the perfect-rationality paradigm is the assumption of perfectly rational opponents. As Aumann (1997) argues, people and organisations use "rules of

[2] Some of these algorithms may converge to the best strategy only in the limit, when an agent plays against a stationary environment. Still, in multi-agent systems where several players learn and modify their behaviour at the same time, there is no proof of convergence to the best strategy.

thumb", learned from experience, when acting, and therefore there is a need to model an opponent's behaviour, i.e., a decision-maker faces the problem of finding and exploiting the weaknesses of his opponents.

In order to apply this bounded rationality concept to the analysis of human behaviour a computational framework that can model decentralised decision-making is agent-based simulation. Further, it can be used in complex systems, being able to model several levels of rationality and different architectural structures. Finally, it may capture learning in games.

1.2. A Motivation from the England and Wales Electricity Market

During the 90s, the privatisation of the E&W electricity industry (more specifically generators, suppliers and regional electricity companies) implied a very important structural change[3]. This privatisation process aimed to introduce competition in the generation and supply businesses, while maintaining a monopoly regime in the transmission and distribution.

Briefly, the first years of the liberalised electricity industry can, in general terms, be characterised as follows.

Before privatisation, the Central Electricity Generating Board (CEGB) dominated the industry structure, selling electricity in bulk to 12 area distribution boards, each of which served a closed supply area. This monopolistic system was characterised by centrally planned investment and an engineering approach to the management of electricity plants.

[3] A good source of information on the history of the electricity sector in the UK is the Electricity Association: A brief history of the UK electricity sector (Electricity Association, 2000a).

The 1988 White Paper (Cm322 Privatising Electricity) set the foundations of this liberalisation process. The 1989 Electricity Act specified how the actual privatisation would take place. It implied four main developments: a change of ownership from the state to private owners; a transfer of employees to the successor company; an introduction of competitive markets and a system of independent regulation.

The liberalisation started by splitting the CEGB into three different generation companies (National Power, PowerGen and Nuclear Electric) and a transmission company (National Grid Company). The policy-makers in England and Wales assigned the fossil-fuel plants to National Power (38 GW of capacity) and PowerGen (18 GW of capacity), and assigned the nuclear plants to Nuclear Electric (8.4 GW of capacity). Hence, in 1992, the first *tranche* of National Power and PowerGen (60% of each company) was privatised; the second *tranche* (40%) was privatised in 1995. In 1996, the policy-makers split Nuclear Electric into two companies, Magnox Electric and British Energy (who received the more advanced nuclear power stations, the gas-cooled reactors).

It was then that the restructuring process entered a second stage with successive divestments and the arrival of new players to the E&W market. Thus, in 1995, Edison Mission Energy entered the market buying the pumped storage business. In 1996, TXU Europe (former Easter Group) entered the market by acquiring five coal-fired power stations, under a 99-year lease from National Power and PowerGen, with a total capacity of 6 GW. At the same time, the Scottish companies and Electricité de France supply power via interconnectors participating actively in the E&W electricity market. This restructuring process continued and, in early 2000, there were 24 companies in the market: at this time, the small Independent Power Producers (IPPs) had 21% of the installed capacity.

However, before proceeding it is noteworthy that the arrival of the combined cycle gas turbines (CCGTs) facilitated this liberalisation process. This technology reduced the economies of scale in electricity generation, allowing greater flexibility in investment decisions (due to low capital cost and short construction time).

Consequently, the installation of new players, vertical integration (between generation, distribution and supply), and diversification (the main electricity companies are also selling gas) reshaped the industry. From this evolutionary process, two main features emerged in these first years of the E&W liberalised electricity market. The first was the plant trading among generators. The incumbents divested under regulatory pressure, while the new players built and bought plants to enter the market. The second was the greater pool price volatility; Bunn (1997: 2) argues that one of the sources of price uncertainty was generators intervention in the Pool[4]. Further, Bunn (1999: 3) argues that as the regulator controlled average prices the "Volatility would clearly encourage contracting and risk premiums to the generator's benefit…."

The most recent development in the E&W electricity market was the introduction of NETA (which occurred in March 2001). NETA replaced the mandatory daily uniform price auction, aiming at improving the participation of the demand side in the wholesale electricity market and at rewarding flexible plants.

All these events seem to suggest two possible issues worth researching:

1. To develop an agent-based computational model of NETA. This aims to evaluate the impact of NETA on companies' performance and to clarify the relation between bilateral markets and the balancing mechanism (the two "market places" where the companies could trade electricity).

2. To analyse the structural evolution of the E&W electricity market, specifically focusing on the strategic management of plant portfolios.

[4] Three references about the Pool are Green (1996, 1999) and Lowrey (1997).

1.3. Overview of Thesis

The main contributions of this thesis are:

1. Improved agent learning algorithms. The thesis presents principles and general methods for the development of learning and decision-making algorithms for interdependent agents in market environments with interdependence. The methods are well grounded in behavioural game theory and automata theory, and are intended to produce sensible (but not necessarily payoff-maximizing, nor supportive of pareto-dominant equilibria) economic behaviour under different institutional conditions. Most importantly, the approaches endow agents with prior beliefs or models, and work without the vast numbers of repetitions and feedback usually required by learning algorithms. The principles are primarily developed in Chapters 3 and 6.

2. Credible, institutionally realistic simulations. In the thesis, practical and insightful simulation models are developed and successfully run based on the applications of the principles from computational learning theory to several market environments specific to the electric power industry. Agents in the model use decision-making and learning algorithms developed on behalf of electricity generators and suppliers. Endogenous to the models are important variables such as pricing levels, capacity available, firms' plant portfolios, and inter-firm plant transactions. When "let loose" in detailed and institutionally realistic computer-simulated market settings, the agents and their behaviour provide insights into industry structure, conduct, and performance characteristics. Successful applications of agent algorithms are demonstrated in Chapters 4 and 5.

3. Development of a theoretical framework for simulation of market structure dynamics. Specifically, Chapter 7 uses agent-based algorithms for the

analysis of the evolutionary properties of electricity markets. This represents another way to use simulation to develop theoretical insight into real world problems. By developing an analytical and algorithmic framework for the analysis of electricity markets short-term dynamics the use of simulation allows the analysis of different scenarios which do not intend to replicate reality but are aimed at a theoretical analysis of electricity markets' behaviour. Hence, in this setting, the mix of a behaviourally grounded theoretical model with the potential of simulation analysis gives an insight into how players value their plant, into the reasons why they trade plant and, most importantly, gives insight into the interdependence between market rules, technological constraints, and the evolutionary properties of electricity markets.

4. Objective platform for policy analysis. The agent algorithm and market simulations provide an objective platform for policy analysis in the electricity industry. Simple, self-interested agent behaviour (in the absence of collusion or cartel agreements) is shown to lead to capacity withholding, price spikes, strategic groups, and active secondary markets for generation facilities. While this thesis is not intended to present a policy analysis, the methods developed and demonstrated are shown to have the potential to inform policy discussions and provide an understanding of how industry dynamics arise from the behaviour of profit-motivated agents under conditions of strategic dependence and uncertainty.

Next, Chapter 2 presents a literature review of electricity markets addressing the issues of market power, market design and trading of plants. By looking at each of these specific research questions, this literature review aims to emphasise how each one of them posed methodological challenges to the electricity markets' research community, and to describe how several methods have been dealing with these challenges. Furthermore, this review also looks at the historical evolution of the research on electricity markets, pointing out the limitations of the several methods used and identifying several still-open theoretical questions, some of which are addressed in this thesis.

Having reviewed the literature on the modelling of electricity markets, this thesis proceeds by addressing the development of computational models of learning within games. Specifically, Chapter 3 motivates the use of learning in games, discusses the advantages and drawbacks of reinforcement learning and introduces the need to model opponents' behaviour.

Having discussed the political and methodological issues addressed in this thesis, this analysis proceeds in Chapter 4 by presenting an application of agent-based computational learning to simulate the New Electricity Trading Arrangements in the E&W electricity market (before its introduction in 2001). This work develops a large-scale application of multi-agent evolutionary modelling to the proposed NETA. This is a detailed plant-by-plant model that at the same time specifies an active demand side. This model simulates the interaction of generation and supplier companies, under NETA, analysing the effect of capacity margin on prices and the strategic implications of the new market, both for generation and supply companies, and showing that agent-based computational methods can provide insights into pricing and strategic behaviour in complex new markets such as electricity.

Chapter 5 presents an analysis of the learning processes and pricing policies affecting the potential for the exercise of market power. Based on the simulations carried out in the NETA simulation platform, this analysis aims to explain how individual companies can profit from price manipulation in the E&W liberalised electricity market. The specific modelling challenge is therefore to assess, *ex-ante*, if such players can really influence market prices profitably by attempting strategic capacity withholding. Further, Chapter 5 presents a simple theoretical contribution by extending Crampes and Creti's (2001, October) analysis of simple Bertrand games to an N-player pay-as-bid game.

Hence, having addressed some important policy issues in market design and market power analysis by using an *n-armed bandit* model of learning, the thesis then prepares the analysis of structural evolution in electricity markets and more specifically plant trading. This issue represents a new challenge for computational learning as the number of interactions with other players (and therefore opportunities to learn), when strategic actions are concerned, is extremely limited, and also as some of these actions may be extremely costly and irreversible. This plant trading issue motivates the development of the Finite Automata Dynamic Game (FADG).

Chapter 6 develops the FADG, studying its theoretical properties. In an FADG, the players (modelled as finite automata) learn rationally and adapt as the game proceeds, developing a forward-looking behaviour, even when only a few interactions with the environment are possible. The FADG models the set of possible automata as endogenous variables, capturing the process by which a model of other players' behaviour is learned, and enabling the analysis of the processes by which certain types of behaviour emerge. Further, the analysis of the FADG shows that best response is not a sufficient condition for rational behaviour in a world where the presence of boundedly rational players is the main factor creating uncertainty.

Thus, having developed a game that is able to deal with problems where actions may be irreversible (and costly), and (or) players seldom interact, it is now possible to address the plant trading issue. Chapter 7 starts by developing a theoretical model to justify plant trading in electricity markets and to analyse the main drivers of short-term market structure dynamics. The analysis of this model shows that, within the conditions defined by the model, capacity withholding is the main driver of plant trading, and that this activity leads to higher market concentration.

Further, Chapter 7 presents a dynamic game to analyse how players learn and adapt in order to coordinate the actions needed for plant trading, and in order to develop a model for plant valuation. Moreover, as by definition all the trading takes place in off-equilibrium states of the game, it follows that in a trading game the trajectory toward equilibrium is as important, if not more important, than the equilibrium itself. Thus, this chapter concludes by emphasising the need to model learning and adaptation, and to use the FADG, in games with the characteristics of the plant trading game.

Finally, in Chapter 8 the main contributions of this thesis are summarised.

CHAPTER 2

MODELLING ELECTRICITY MARKETS

The liberalisation of electricity markets is continuously challenging the modelling techniques available for market design and analysis of evolutionary systems. This chapter summarises the research addressing the topic of liberalised electricity markets, emphasising the way new methodologies are required to accommodate its demands.

The recent literature on modelling of electricity markets addresses three main topics:

1. Market Power and Market Design. This has been the main topic of research as policy makers, companies and academics are looking at the best way to organise trading activity in liberalised electricity markets. Further, the work on conduct-evaluation addresses the topic of market power from a behavioural perspective (both within modelling and empirical studies), without directly referring to market design.

2. Strategic Management of Plant Portfolios, more specifically, the problem of investment planning and regulation in the liberalised markets, analysing how the liberalisation process affects investment decisions in generation.

3. Economic Dispatch and Unit commitment. This is the problem faced by a generator, in a liberalised market, that specifies how to price his plants in order to maximise the profit of his portfolio.

Next, section 2.1 discusses the topic of market power and market design, and section 2.2 analyses strategic management of plant portfolios; finally, section 2.3 gives a brief summary of the unit commitment problem in decentralised markets.

2.1 Market Power and Market Design

During the 90s, the issue of market design has been the main concern of policy makers and academics. Privatisation (aiming to introduce competition in the generation and supply businesses) changed the industry ownership structure by developing a competitive wholesale electricity market, and a system of independent regulation, while maintaining a monopoly regime in the transmission and distribution businesses.

The main research question addressed in the literature during those years was the definition of the trading mechanisms and market regulations that would minimise the exercise of market power by generators. Further, the central tool used to analyse the market design issue, specifically in the Pool type of markets, was the supply function equilibrium developed by Klemperer and Meyer (1989). The model of supply function equilibrium captures an oligopoly wherein each firm chooses to bid a supply function relating prices and quantities. This is widely accepted as the most adequate tool to model an electricity market such as the E&W Pool. In this market, each generation company had to submit bids for each one of their electricity plants using a kinked supply function. Theoretically, the supply function equilibrium (a pure strategy Nash equilibrium) builds a bridge between the Cournot and the Bertrand models. The steeper the supply function the more the equilibrium approaches the Cournot model (in which quantity is the strategic variable). The flatter the supply function the closer the equilibrium approaches the Bertrand solution (in which price is the strategic variable). Another advantage of supply function model is to enable a more flexible "strategy" (and a higher expected profit) for firms that do not commit to a fixed quantity or price. However, although computing the supply function equilibrium this model does not explain how to reach it.

Green and Newbery (1992) develop a stylised oligopoly model of competition using the supply function equilibrium, showing that, at that time, there was a high possibility for generators to exercise market power in the E&W electricity market. They also suggest that divestment was needed in order to get a more competitive market, showing that the two main generators at the time, National Power and PowerGen, still had high market power, even without collusion, offering their capacity at prices above marginal costs, and exploiting the constraints on the transmission capacities. Nonetheless, Green and Newbery (1992) also argue that, in the long run, a high supply low-price strategy, avoiding new entrants, would be the best choice for the incumbent firms. Further, Green (1996) analyses again the E&W wholesale electricity market advising three possible policies to reduce the generators' market power. First, he suggests divestment as a policy to force incumbent companies to sell some of their stations. Second, he suggests demerging or breaking-up of the incumbents into a number of smaller companies. Finally, he emphasises the need to encourage new entry.

Thus, during the Pool years, the market developed a hedge to the pool auction, the contracts market. Green (1999) adapts the supply function model to analyse this new market, defining a two-stage duopoly game with a spot and a contracts market. He shows that these contracts reduce the incentives for generators to increase the spot price; moreover, since most of the trading occurs in the contracts market, and if the buyers (retailing companies and big consumers) are risk averse, the "contracts market" price may exceed the spot price.

Even so, the generators found ways to manipulate the system to their advantage. Bunn (1997) argues that the generators completely dominated the Pool system and, therefore, that they had the capability to increase spot price volatility, thereby increasing the value of the contracts market.

Additionally, another research topic within market design (and quite an important one) relates the interactions between the different markets within the electricity industry. Herguera (2000) analyses empirically the implications of introducing bilateral markets (namely the contracts for differences and the electricity forward agreements) on the efficiency of the E&W and NordPool electricity pools. He finds evidence supporting the hypothesis (Allaz, 1992; Allaz and Vila, 1993) that, given a concentrated generation market structure, the introduction of a competitive futures or contracts market leads to lower spot equilibrium prices. This hypothesis supports the premise that the introduction of bilateral markets induces generation companies to bid very aggressively in the spot markets (since they secure their position in the bilateral market, where they sell most of their production). This behaviour would drive spot prices close to marginal cost. Herguera finds evidence that the price in the E&W electricity Pool is systematically lower than the price in the bilateral contracts (with the contracts for differences presenting the highest prices). Moreover, he also finds evidence that the amount of production covered by bilateral contracts has decreased in E&W between 1997-1999; this behaviour may have led to the increased pool price volatility. He argues that generators used two strategies to exercise market power during the pool years: price spikes and capacity withholding. Moreover, he maintains the hypothesis that price coordination among generators explains the frequency of capacity withdrawals and price spikes.

The impact of regulatory activity, namely through divestment, is analysed by Day and Bunn (2001) using a computational model, based on the supply function equilibrium, to determine competition outcomes in oligopolistic electricity markets. They simulate the supply function competition, using a best response framework, and assuming kinked, discontinuous supply functions, and compare their results with the theoretical supply function equilibrium. They report finding no convergence to a unique supply function equilibrium: quite the opposite, finding a cyclic behaviour. Moreover, their results seem to support the contention that the 1999 divesture of 40% joint capacity of National Power and PowerGen would not be sufficient to completely erode the possibility of market power exercise in the E&W Pool.

Further, in analysing the California market, Borenstein and Bushnell (1999) suggest that divestment leads to markets that are more competitive. However, it is still not clear that divestment has only positive aspects. On this subject, Ishii and Yan (2002) present evidence that divestment *crowds out* new investment in this market.

In addition to the market design issue, horizontal market power has been analysed *per se*: it has been shown that even in the liberalised markets, generation companies have the capability of manipulating wholesale electricity prices through capacity withholding (Borenstein et al., 1995; Wolak and Patrick, 1997, February; Joskow and Khan, 2002) and high mark-ups (Wolfram, 1998).

Overall, in the first decade of privatised electricity markets, the main concern of regulators has been horizontal market power. Sweetser (1998) identifies several factors that may increase horizontal market power in electricity generation, specifically market concentration, geographical distribution of demand and generation, and the transmission paths between them. Further, on this topic, Borenstein et al. (1995) relate the nature of several electricity-based products with potential exercise of market power. They specifically identify three products. Spot electricity is the first product analysed, in particular a market for baseload and another one for peak demand. Second, they look at pool based and physical power contracts. Finally, they study the reliability services, looking at load balancing, spinning reserve and voltage support. They emphasise the role of transmission congestion in splitting markets as an important factor to take into account when analysing market power in the peak demand hours. Furthermore, they support the existence of a market for contracts that includes contracts for differences and direct contracts on electricity supply, and contracts for transmission congestion. In addition, they advise the creation of a market for ancillary grid and reliability services.

Another central issue on the topic of market power is to find a measure to detect the existence of conditions for abuse of dominant position. Borenstein et al. (1999) maintain that traditional concentration measures, such as the Herfindahl-Hirschman Index, are likely to be inadequate for this task, identifying three main factors (besides industry concentration) that may influence the capability of firms to exercise market power. The first is the structure of incentives for generators. The second is demand elasticity. The third is the potential for capacity expansion and for the arrival of new competitors. Furthermore, they also argue that the features of generation technology (such as ramping rates, start-ups and marginal costs), and capacity and transmission constraints, influence the generation companies' market power. Nonetheless, concentration measures do not consider these factors. Thus, they propose the use of simulation as a tool to capture the strategic behaviour of firms, and their capability to exercise market power, using a Cournot oligopoly with a competitive fringe. They report the existence of multi-equilibria and the potential for abuse of market power in the peak-demand hours (even in a market that seemed *unconcentrated* when evaluated using concentration measures) the reason for which seems to be the low short-run demand elasticity. Further, they argue that incentives to consumers, stimulating higher demand response to price changes, may be more effective in fighting market power than investment in generation and (or) transmission expansion. Finally, they report that transmission constraints increase the possibility of market power abuse.

If the evaluation of market power has been problematic, then equally controversial has been its interaction with the regulatory activity. For example, Garcia et al. (2001) develop a dynamic oligopoly model to capture the interaction between hydro-based and thermal generation in duopoly competition, taking into account the effects of regulatory intervention on the players' behaviour. They show that price caps may influence pricing behaviour even when the equilibrium price is below the price cap. Further, they explain that when hydro generation competes with thermal generation, low price caps (to sell in the winter peak) may lead to system reliability problems, because the hydro generators may decide to empty the reservoirs during the summer (hoping for heavy winter rain). This effect would not occur in the absence of a price cap, as the winter price may be high enough to induce hydro-generators to save capacity for winter. The strategic behaviour, caused by the wrong regulatory incentives, may compromise system reliability. This exemplifies how a regulatory action that intends to avoid peak prices (the price cap) may have unwanted consequences.

On the same subject, von der Fehr and Harbold (1993) model the strategic pricing behaviour of vertically integrated generation players in an environment where there is a price cap and players compete by price in a simple Bertrand game with capacity constraints.

However, none of the modelling techniques referred to above (supply functions equilibrium, Cournot or Bertrand models) capture the adaptation process and the interaction possibilities arising in electricity pricing models. Nevertheless, as Rothkopf (1999) reminds us, the interactions within an electricity market constitute a repeated game where a process of experimentation and learning shapes the players' behaviour.

Agent-based simulation is a computational technique that can take these learning processes into account and that can model (with great detail) the interaction between players in the industry. Within this context, the introduction of NETA represented a unique opportunity to test the agent-based simulation in market design. Bower and Bunn (2000) analyse the impact of the introduction of discriminatory pricing on spot market prices under NETA[5]. Their model enabled a very detailed description of the market, taking into account discrete supply functions, adopting different marginal costs for each technology, and modelling the interactions between players in a repeated game.

2.2 Strategic Management of Plant Portfolios

Mergers, acquisitions and divestures are the main problems faced by a regulator and the companies operating in a market (Cox, 1999). Electricity companies may use mergers and acquisitions to adapt to the new environment (e.g. risk management) or to gain market power, whereas divestments by incumbent generation companies may be required in order to ensure that the market is competitive. Therefore, one of the aims of strategic management of plant portfolios may be to gain a dominant position.

Further, market design influences investment decisions. For example, Exelby and Lucas (1993) examine the link between capacity payments and capacity investment in the E&W Pool, showing that the capacity payments mechanism introduced incentives to reduce the capacity available in the system. Moreover, they reveal that the incumbent's optimum behaviour was to adopt a quasimonopoly strategy, and that capacity payments were only an incentive for new entrants to invest.

[5] See Abbink et al. (2003) and Rassenti et al. (2003) for experimental studies in discriminatory pricing.

However, this section, abstracting from the market power and market design issues, analyses the impact of decentralised decision making on strategic management of plant portfolios.

In the E&W electricity market, before privatisation, national security and cost minimisation were the driving forces behind strategic management of plant portfolios. The main goal of this strategy was to assure generation technology diversity. Stirling (1994) claims that technological diversity was the basic rationale for the investment in the UK electricity market, under central planning. Diversification seemed the correct reaction to the uncertainty underlying fuel prices, environmental impacts and financial performance. Sterling also maintains that, since privatisation, the government created better conditions for technological diversity by revoking, in the late 80s, the European Community legislation forbidding the use of gas in bulk power generation. Furthermore, the need for diversification also justified the investment on nuclear plants. Moreover, he argues that in the liberalised market, the new private companies also have technological diversity as one of the drivers of their behaviour as they have indeed diversified their sources of fuel.

Regarding the impact of governmental intervention on the technological diversity of electricity generation in the UK, Henney (1994) suggests that, during the 70s and 80s, the policies were favourable to nuclear and coal technologies. He argues that the UK government knew that nuclear plants had higher unit costs than coal, but they considered the nuclear option an issue of national security. Further, he defends the government support of the coal industry, which they regarded as the main source of energy. In contrast, Newbery (1998) suggests that the UK government promoted the nuclear option as a means to combat coal miners' power. In Henney's view, however, the protection of UK coal led to high inefficiencies and mistakes, namely to spending resources in opening comparatively unproductive mines and over-estimating demand for coal. Further, he says that this bad performance of the coal industry was one of the key factors behind the introduction of gas-fired technology.

Therefore, it seems that under the monopolistic electricity market the main concern of the government, regarding the issue of strategic management of plant portfolios, was technological diversity (that can be considered another facet of national security). The political context of the oil crises in the 70s and instability in the oil producing countries explain this preference for diversification. However, government policy does not explain everything: technological innovation and social development are responsible for an important part of this evolutionary process.

Several factors conditioning the evolution of this industry are now analysed. The driving forces in the electricity industry, during the 80s and beginning of the 90s, were the arrival of new technologies, environmental concerns and political developments (Flavin and Lensen, 1994). The arrival of natural gas allowed independent power producers to build small and efficient power plants. Bodde (1998) identifies some other driving forces shaping the electricity industry structure: *Environmental support*, i.e., there is a need for clean energies (providing a strong motivation for nuclear and gas fired plants). *Dual use technology*, i.e., the nuclear technology can be used to produce nuclear weapons (he also claims that power reactors can make use of plutonium resulting from an arms reduction programme). *Ideology*, i.e., the opposition to nuclear plants is ideological; the industry and governments will need to win the battle of ideas if the nuclear industry wants to survive. *Market acceptability*, i.e., the decisions on adoption of new technologies will tend to be economic. Prices seem to be at the heart of competition, while product differentiation appeared extremely difficult to achieve. Furthermore, Bodde identifies several possible sources of business risk (the large investment required, the fuel prices variability, and the exogenous safety risk of nuclear plants), concluding that the evolution of competitive markets would not favour nuclear technology.

However, the logic behind the strategic management of plant portfolios in electricity markets has undergone a major shift with the privatisation and deregulation processes. Larsen and Bunn (1999) summarise the changes in the industry resulting from the privatisation process in the E&W electricity market. At an industrial level, the new industry is characterised by unstable and volatile prices, the presence of new shareholders with diverse objectives, regulatory uncertainty, and information opacity (the price signalling effects may be misleading for investment). At the corporate level, the new market is characterised by a focus on shareholder value (that replaces the social optimum) and new methods of linking strategic thinking, uncertainty, and limited information (replacing the classic operational research planning). Further, Larsen and Bunn (1999) identify three types of risk that may affect companies in the new oligopolistic markets. The first is corporate risk that they define as cultural change in order to adapt to the new market. The second is market risk that is due to price volatility and strategic interaction with other players. The third is regulatory risk, specifically the way a regulator chooses to balance controls on prices and investment policies, defending the consumers' interest and maintaining the industry attractiveness. Hence, it is clear that regulatory choices may have important impacts on shareholder value.

Within the process of liberalisation, another central issue related to the strategic management of plant portfolios is the type of privatisation to adopt, i.e., diversification versus vertical integration.

This topic is addressed by Kaserman and Mayo (1991) who claim that the industry should be privatised vertically due to the presence of economies of vertical integration, and due to the exhaustion of economies of scale (caused by technological change and by electricity demand growth). It is noteworthy that the evolution of the E&W electricity market seems to support their hypothesis. Even though privatised horizontally, ten years after privatisation, the E&W electricity industry has converged towards vertical integration. In addition, on this topic, Kennedy (1997) analyses the way vertical integration affects market power, wholesale prices, and barriers to entry. Kennedy argues that vertical integration benefits would depend on the market structure (if the supply is regulated and there is competition on the generation side then vertical integration reduces transaction costs). Nonetheless, he identifies vertical integration as a possible barrier to entry, as the merging of supply and incumbent generation may work as a threat for potential new entrants.

Therefore, it seems that if technological evolution was essential for liberalisation to be possible, the political will for privatisation was equally essential for it to happen. Nevertheless, economics still plays a role in the genesis of these new markets, as the issues of market design and market power, together with the implications of vertical integration and diversification, are ever present.

Nonetheless, if the application of game theory and agent-based models was the main methodological transformation in research on electricity markets, strategic management of plant portfolios also brings some new challenges to research methods.

Within the new liberalised markets, and due to the decentralisation of the long-term decisions, the investment problem became an important research issue. The privatised market presents an increased risk due to price and demand uncertainty and due to competition (the investment projects are private). Further, technological innovation is unpredictable and some baseload investments, such as nuclear or hydro stations, take a long time to complete. *Thus, there is a need to explain how electricity companies and these markets evolve in the short and long-term, and to explain the impact of uncertainty on the value of electricity plants.*

Skantze et al. (2000) address the investment problem in oligopolistic electricity markets with a two stage dynamic programming approach, with stochastic prices (for both electricity and fuel), to perform simulation-based valuation of generation assets, taking into account start-up and shutdown costs. Within the same topic, Visudhiphan et al. (2001) model investment dynamics in a system with a spot and futures market, analysing how price information affects long-term supply, demand and price evolution. In this model, demand and generation evolution is not explained by optimising behaviour, and market structure is not considered a determinant of generation behaviour and price evolution, nor does it take into account any issues regarding vertical integration between generation and supply. The model assumes that electricity prices follow a stochastic process depending on aggregate demand and generation evolution.

Visudhiphan et al. claim that the introduction of a market for futures stimulates investment, because it anticipates future incomes for generation companies. They simulate investment behaviour using a backward-looking strategy, wherein investment depends on past spot prices, and a forward-looking strategy in which investment depends on the prices in the market for futures. Their simulations show that a backward-looking strategy leads to investment delays and under-investment, while a forward-looking strategy leads to smaller imbalances between generation and demand.

Pineau and Murto (2003) put forward an interesting approach to model investment behaviour by using a stochastic Cournot oligopoly model for the analysis of the Finnish electricity market. Within this model, they capture demand as a stochastic process and define generation capacity as a function of players' strategies and stochastic elements. Further, generators can only invest in gas and coal plants (the period modelled was 10 years) and the market is segmented into baseload and peak demand. The alarming conclusion of this work is that the higher uncertainty in the liberalised electricity markets, when compared with the monopoly situation, and in the presence of big players, leads to lower investments and, in the long-run, may threaten system reliability and lead to high prices.

2.3 Unit-Commitment Problem

It is noteworthy the revival of the literature on the unit commitment problem, in an attempt to adapt the classic models (e.g., Bunn and Paschentis, 1983), to the new restructured electricity markets.

Takriti et al. (2000) develop a stochastic model for the unit commitment problem incorporating electricity trading, and stochastic prices and loads. Unfortunately, this model has two main drawbacks that make it inappropriate for modelling investment planning: first, it disregards the strategic interactions between players (a generator is assumed to be a price taker); second, the curse of dimensionality, it only models a typical week.

Further, Anderson and Philpott (2002a, 2002b, 2003) address the unit-commitment problem within an electricity pool. Their work can be classified into two categories (control and forecasting), which correspond to the basic problems faced by a player in a dynamic system (as further analysed in Chapter 3).

The control problem is addressed by Anderson and Philpott (2002a) who analyse how a generator can compute an optimal supply function, for a certain market distribution and assuming constant the opponents' behaviour. Further, Anderson and Philpott (2002b) extend the supply function equilibrium with n symmetric generators (in piecewise linear convex cost functions), proved by Rudkevich et al. (1998), to the case of general convex costs (assuming identical generators).

Anderson and Philpott (2003) address the forecasting problem looking at the process by which a generator infers a market distribution for a given time, assuming Bayesian inference and that the opponents of the player in question use a constant supply function, and where uncertainty arises from demand.

Finally, it seems that the unit commitment and market power problems are two faces of the same coin. While the former represents a normative approach to profit maximisation, the latter corresponds to a positive one. The first computes the best way for a generator to increase his profit while the second considers whether this behaviour constitutes abuse of a dominant position.

2.4 Summary

The liberalised electricity markets stretched the existing modelling techniques to their limit. There is a need to develop modelling techniques that are able to capture, at the same time, the interaction between companies and the internal processes within them. These goals may be achieved using the agent-based computational learning models presented in Chapter 3, as this technique can handle, simultaneously, a detailed description of the players in the model (including their adaptive behaviour) and the interactions between them.

Further, as presented in this chapter, different researchers have different reasons to use different models.

Thus, researchers focusing on the non-linearity of competition and on price competition between players tend to prefer the Bertrand model. See Staiger and Wolak (1992), von der Fehr and Harbord (1993), and Crampes and Creti (2001, October), for Bertrand games in the electricity industry, and Kreps and Scheinkman (1983) for a game theory analysis. Further, Bower and Bunn (2000) present a market design application of an agent-based model where players use prices as instrument.

However, the Bertrand game with capacity constraints is hard to solve (it is a non-linear optimisation problem). Moreover, section 5.1 shows that even in very simple settings this model may have no equilibrium.

On the other hand, the supply function game has been the favourite of researchers analysing the behaviour in pool markets where generators offer a supply function (Green and Newbery, 1992; Day and Bunn, 2001; Anderson and Philpott, 2002a, 2002b, 2003). The research on this model has improved its ability to deal with some of the complexities of discontinuous functions. However, it is difficult to envision its applicability in liberalised markets organised around bilateral trading.

Finally, the Cournot model has been extensively used in the electricity market's game-theoretical literature. Allaz and Vila (1993) analyse Cournot competition in forward markets. Borenstein and Bushnell (1999) use it to analyse market power and divestment in the California electricity market. Wei and Smeers (1999) and Hobbs (2001) analyse spatial competition in restructured electricity markets assuming Cournot behaviour. Finally, Pineau and Murto (2003) look at the investment problem using a Cournot model.

Hence, the modelling of liberalised electricity markets requires computational techniques that bring together the detail and computer power of Operational Research with the behavioural analysis of Economics.

At a theoretical level, it would seem that the main topics presented still have many unanswered questions.

Within the market power and market design topic, the evaluation of the ability of a player to exercise market power is an open question (Chapter 5 addresses this topic). Further, the interaction between suppliers and generators is another important and challenging issue, which this thesis addresses in Chapter 4, using the simulation of NETA.

However, due to information scarcity, the strategic management of plant portfolios is the least researched topic. As referred to above, a new theory is required to explain market structure evolution, and to analyse the impact of uncertainty and learning processes on the value of electricity plants. Chapter 7 addresses this problem by analysing the issue of plant trading in liberalised electricity markets.

CHAPTER 3

MODELLING LEARNING IN GAMES

In games with multiple equilibria, the perfect-rationality paradigm, i.e., Nash equilibrium, fails to predict the players' behaviour. Empirical studies (e.g., Roth and Erev, 1995) show that models of bounded rationality predict better than the Nash equilibrium how people, organisations and markets behave (at least in the short run). Therefore, in order to model complex games possibly with multiple equilibria, computer models that incorporate boundedly rational players are used as a mechanism for inductive equilibrium selection, and to test the validity of the perfect-rationality predictions.

One model that captures boundedly rational behaviour is reinforcement learning (Sutton, 1988; Weiss, 1995; Bertsekas and Tsitsiklis, 1996). By using an algorithm of reinforcement learning, an agent plays a game in the extensive form with incomplete information where, at the end of each stage, the player only observes the outcome of the game (the transition between states) and the reward he receives.

This chapter analyses the use of computational learning in agent-based models. Why are models with learning agents useful? There are several types of agent-based models, some of which do *not* consider learning, focusing on the interactions between players (the emergent behaviour). These models (such as cellular automata) are based on simple agents that always react in the same way to the same stimulus. Applications of models based on cellular automata have been used both in organisational modelling (Lomi and Larsen, 1996) and, with great success, in geographic and urban economics (Ward et. al., 2000). Other types of agent-based models use agents that are more sophisticated, having learning and reasoning capabilities. This is the type of model analysed in this thesis.

The theory of learning in games assumes that agents learn by interacting with each other. More importantly, agents tend to repeat the actions that give them the best rewards, i.e., they learn by reinforcement of positive actions.

However, this methodological jump from perfect-rationality to bounded rationality has theoretical and philosophical implications. It corresponds to a switch from a "normative theory" to a "positive theory." The normative theory prescribes what each player in a game should do in order to promote his interests optimally (von Neumann and Morgenstern, 1953; van Damme, 1991: 1), while the positive theory describes how agents actually decide. This line of research tries to understand how people and institutions behave (Samuelson, 1997: 3).

Although having had success in the analysis of strategic behaviour, the positive approach to game theory has faced difficulties to prescribe a "rational behaviour" in games exhibiting multi-equilibria. A first attempt from the game theory literature to address this issue was to refine the concept of Nash equilibrium by including additional criteria. First, a player does not choose dominated strategies (Fudenberg and Tirole, 1991: 8). Second, choices in information sets not in the equilibrium path must be optimal choices (in order to avoid non-credible threats). This is called the Rationalizability criterion (Bernheim, 1984; and Pearce, 1984).

However, the problem of equilibria selection still exists as different refinements select different equilibria. Further, rationalizable strategies may be too demanding as they assume common knowledge of rationality. In some games, a player does not know if his opponents are rational; besides, there is no common knowledge of rationality. Moreover, as Samuelson points out, in normative game theory common knowledge of rationality is a necessary condition for the existence of equilibrium.

In addition, a positive game theory takes into account that models are just simplified representations of the actual games that people play: Selten's (1975) trembling hand perfection assumes that a player chooses the action that maximises his utility most of the time, but sometimes, by mistake, may play something completely different. Selten presents "complete rationality" as a limit case of "incomplete rationality."

However, this issue is not as black and white as Samuelson's analysis suggests. If there is no common knowledge of rationality, before acting a player needs to infer a model of how the others behave. In this case, positive and normative approaches are not separate anymore, as any normative attitude implies an understanding of the opponent's real behaviour.

Simon (1972) was the first to emphasise this need to model bounded rationality in order to capture human and organisational behaviour. As Aumann (1997) explains, people and organisations use "rules of thumb" that they learned from experience when acting. In other words, people do not optimise even in simple decision problems. This argument underlines the need to model the opponent's behaviour, which is formalised by Rubinstein (1986, 1998) using deterministic finite automata (see Hopcroft and Ullman, 1979 for an introduction to automata theory).

Thus, this analysis of the literature suggests that by abandoning the "perfect-rationality paradigm", a positive-normative analysis of human behaviour implies two possible ways forward. The first is to model reinforcement learning, capturing the behaviour of a player who learns to repeat the actions that generate the highest utility. The second is to model games where a player learns a model of his opponents in order to understand the rules of behaviour they use, and then adapts to the model inferred.

As Chapters 4 and 5 apply reinforcement learning to the analysis of the pricing behaviour in the electricity industry, section 3.1 presents this learning algorithm, discussing its advantages and limitations, and section 3.2 introduces the opponent modelling approach.

3.1 The Reinforcement Learning Problem

The reinforcement learning algorithms can approximate the best policy to play within a certain environment without building an explicit model of it. In the first type of model presented (the *n*-armed bandit), an agent decides how to play in the next iteration given the expected utility (profit) of each possible action.

Other reinforcement learning algorithms, such as temporal differences or Q-learning, are more complex. The agent evaluates each action (in the possible states of the world) taking into account its possible repercussions in the future behaviour of the system. These algorithms are very similar to dynamic programming that rational agents use to approximate the best policy when a full model of the environment is available (Bertsekas, 2000).

This chapter summarises the reinforcement learning theory emphasising its relation with dynamic programming (reinforcement learning algorithms may replace dynamic programming when a full model of the environment is *not* available) and analysing its limitations as a tool for modelling human and organisational behaviour (Barto et al., 1995; Sutton and Barto, 1998; and Bertsekas and Tsitsiklis, 1996). Before proceeding, this section presents the terminology used in models of learning, which follows Singh (1994).

The first concept presented is the *Markov property* (Howard, 1971). Let $P\left(s_{t+1} = s' \mid s_t, s_{t-1}, \ldots, s_0\right)$, $\forall s', s_t, s_{t-1}, \ldots, s_0$ represent the probability of transition from a state s_t to a state s', where s_i is the state of a dynamic system at time *i*. If the state has the Markov Property, it contains all the information to determine the transition probabilities and, in this case, the model transition dynamics is $P\left(s_{t+1} = s' \mid s_t\right)$.

If every state has the Markov Property, the environment and the task as a whole also have the Markov Property.

A Markovian decision process is a tuple (*policy*, *reward function*, *value function*, *model of the environment*)[6].

A *policy* $\pi : A(s) \rightarrow a$ is a rule of behaviour that transforms states into actions.

A *reward function* $u_{ij}^{a_t}$ defines the utility that an agent receives from choosing an action a_t in a certain state *i*, at a given time t. The ultimate goal of an agent is to maximise the total reward received in the long-run, his *Return* (R_t). If the horizon is finite with *T* stages, then $R_t = u_{s_t, s_{t+1}}^{a_t} + u_{s_{t+1}, s_{t+2}}^{a_{t+1}} \ldots + u_{s_T, s_{T+1}}^{a_T}$.

If the time horizon is infinite, the return is the discounted sum (where $0 \leq \gamma \leq 1$ is the *discount* parameter) of each one of the rewards $R_t = \sum_{k=0}^{+\infty} \gamma^k u_{t+k+1}$.

[6] In this thesis, only problems with discrete state spaces are analysed. However, reinforcement learning can also be used to solve problems with continuous discrete spaces. However, these algorithms are less efficient and imply the solution of an additional problem known as generalisation (Bradtke and Barto, 1996; Tsitsiklis and Roy, 1997), i.e., to infer the value of a certain state of the world based on other states.

A *value function*, *V(s)*, specifies the value of the state *s*, i.e., the total value of rewards an agent expects to receive until the end of the horizon by entering that state immediately.

When the problem has discrete state spaces and actions, and if the agent has a perfect knowledge of the transition probabilities and rewards, in order to compute the optimal policy there is no need to interact with the environment. The method used to compute the optimal policy is dynamic programming.

A *model* of the environment is the set of state transition probabilities and transition rewards associated with every action in each state. This model is just optional in reinforcement learning but it is essential for dynamic programming. One of the advantages of reinforcement learning is that it enables an agent to have forward-looking behaviour without having an explicit or complete model of the environment. However, an agent is required to interact several times (a high number) with this environment.

An algorithm that works *on-line* learns the value function and controls a real environment at the same time. On-line algorithms present a trade-off between *exploration* (an attempt to improve the knowledge possessed at a certain time) and *exploitation* (an attempt to profit from the knowledge of the environment). In order to increase its knowledge of the environment an agent may have to choose sub-optimal actions. If an agent chooses only optimal actions, he may end up exploiting a local optimum, and may not converge to the *global optimum* solution. *Off-line* algorithms use simulated experience with a model of the environment and do not face the exploitation vs. exploration trade-off. These algorithms learn the control policy before applying it to the real world. Note that the distinction between on-line and off-line algorithms, and real versus simulated world, is somewhat misleading. An algorithm that does not require a model of the environment can solve a problem in which a model is available. In this case, a player uses the model to simulate the real world.

Next, dynamic programming and three main approaches to reinforcement learning are discussed, namely the *n*-armed bandit, the temporal differences and the Q-learning algorithms. As seen in the previous presentation, reinforcement learning and dynamic programming can be used to solve the same type of problems, as both have the purpose of computing the *optimal policies* given that the environment follows a Markov process. Therefore, this chapter describes reinforcement learning as a generic problem of which dynamic programming is a special case: this approach follows Sutton and Barto (1998), Bertsekas and Tsitsiklis (1996) and Kaelbling et al. (1996).

3.1.1 Dynamic Programming

Bellman (1957) defines dynamic programming as the mathematical theory of multi-stage decision processes, and defines a *decision process* as a system (possibly stochastic) in which the decision-maker has a choice of transformations that may be applied to the system at any time. He also distinguishes single-stage (only one decision has to be made) from multi-stage decision processes (where the decision-maker has to take a sequence of decisions). A decision maker solves a dynamic programming problem when he understands the *structure* of the optimal policy. This means that the decision-maker understands the system characteristics, which determine the decisions at any particular stage of the process. An optimal policy has the property that whatever the initial state and initial decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision (this is the *Principle of Optimality*).

Therefore, the dynamic programming problem is the one of computing the *optimal policies*, given that the environment follows a Markovian process and a known model of the environment. Hence, it implies the solution of two related problems: prediction (policy evaluation) and control.

The *prediction* problem consists of estimating the value of each state of the environment for a certain policy π.

Let E_π represent the expected value of a policy π, and let $\pi(s,a)$ represents the probability of executing action *a* in state *s*, following a policy π. Further, let $P^a_{ss'}$ stand for the state transition probabilities, representing the probability of the system moving from state s to state s', conditional on the agent choosing action a.

Then the *state-value function* $V^\pi(s)$ represents the expected return of state *s*, following a policy π such that $V^\pi(s) = E_\pi\{R_t \setminus s_t = s\}$, or equivalently
$V^\pi(s) = \sum_a \pi(s,a) \sum_{s'} P^a_{ss'}\left(u^a_{ss'} + \gamma V^\pi(s')\right)$.

The control problem takes into account that a given policy influences the value of a given action. Thus, let the *action-value function* $Q^\pi(s,a)$ represent the expected return of an action *a,* at state *s,* following policy π. Then, $Q^\pi(s,a) = E_\pi\{R_t \setminus s_t = s, a_t = a\}$ represents the action-value function, which is equivalent to $Q^\pi(s,a) = \sum_{s'} P^a_{ss'}\left(u^a_{ss'} + \gamma V^\pi(s')\right)$.

It is then intuitive that a value function imposes a partial order on the space of policies. A policy π is better than or equal to a policy π' if and only if its expected value is higher than or equal to the expected return of policy π'. More formally,
$\pi \succ \pi' \Leftrightarrow V^\pi(s) \geq V^{\pi'}(s)$, for all s $\in S$.

Since it is possible to have a partial ordering on the space of policies, it is also possible to define optimal value functions. The optimal state-value function

$V^*(s) = \max_{\pi} V^{\pi}(s), \forall s \in S$ represents the maximum expected value of a state s. Therefore, the optimal action-value function $Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a), \forall s \in S, a \in A(s)$ represents the maximum expected value of an action a in a state s.

Hence, an optimal policy π^* gives the association between states of the world and actions that maximise the expected return, $Q^*(s,a)$:

$$\pi^*(s) = \arg\max_{a} \sum_{s'} P^a_{ss'} \left\{ u^a_{ss'} + \gamma V^{\pi}(s') \right\},\ \forall s \in S.$$

For algorithms with an infinite horizon, the value iteration and the policy iteration algorithms, first developed by Howard (1960), can approximate the optimal policy and the optimal value functions. The *value iteration* algorithm is a successive approximation method, in the space of the value functions *V(s)*, that converges to the optimal value function, in the limit. The *policy iteration* algorithm is an approximation method, in the space of policies, which converges to the optimal policy, in the limit (Ross, 1983; Bertsekas, 2000).

3.1.2 *N*-Armed Bandit

In the n-armed bandit problem (see Sutton and Barto, 1998) a player needs to choose between n possible actions. After each action, the player receives a reward. A player's objective is to choose the actions that maximise the overall reward in a certain period.

Usually, this is a stochastic problem (the rewards are generated by a stationary stochastic process) where each action has a certain expected value (the expected value of the rewards received) unknown to the player. The player's task is to find the action with the highest expected reward and to execute it as often as possible. In order to learn the expected reward associated with each action, an agent is required to try non-optimal actions during a few iterations (exploration) instead of repeatedly choosing the action with the highest expected value at a certain time (exploitation).

Let $Q_t(a)$ stand for the expected reward of action *a*, at stage *t*, and let *a* represent one of the possible *n* actions available to the agent. Further, let u_{t+1}^a stand for the reward received by executing action *a* at iteration *t+1*. An agent attempts to estimate the true value of *a*, *Q*(*a*), computing on-line equation (3.1).

$$Q_{t+1}(a) = Q_t(a) + \alpha.\left[u_{t+1}^a - Q_t(a)\right], \ \forall a. \tag{3.1}$$

Equation (3.1) represents an exponential smoothing of past rewards with a weight-factor α, also known as learning rate, such that $0 \le \alpha \le 1$.

Nevertheless, this learning equation carries no guaranteed convergence properties. The learning algorithm converges with probability one for the true action value only if the learning parameter decays over time, let us say $\alpha_t = \frac{1}{t}$. However, only in games where rewards are stationary is this condition valid. The conditions for convergence with probability one are presented in equation (3.2), i.e., the learning rate converges to zero but slowly, meaning that the agent can learn fast enough to revise his initial expectations and slowly enough not to incorporate the effects of random events into his expectations.

$$\sum_{t=1}^{+\infty} \alpha_t(a) = +\infty \text{ and } \sum_{t=1}^{+\infty} \alpha_t^2(a) < +\infty . \tag{3.2}$$

In economics, Arthur (1991) using this type of algorithm to capture bounded rationality finds that computer automata (after calibration) learned in a very similar way to humans, in the sense that they also deviate from perfect-rationality. Roth and Erev (1995) use an n-armed bandit algorithm to capture the effect of experience and learning in human behaviour they show that learning models predict better than the Nash equilibrium how humans behave (Feltovitch, 2000; Sarin and Vahid, 2001). They also show that this learning model fails to converge to perfect equilibrium in the "ultimatum game" (Binmore et al. (1985) observe this same result in their experiments with people).

Finally, an n-armed bandit algorithm has also been used to model electricity markets. Nicolaisen et al. (2001) simulate a wholesale electricity market model using agent-based computational learning. This model aims at analysing the market power issue in electricity markets simulating the interaction between different generation companies, and capturing the players' learning behaviour by using Roth and Erev's (1995) reinforcement learning algorithm.

Chapter 4 presents an application of a simple n-armed bandit algorithm to capture learning and adaptation in a repeated game in an agent-based model of the New Electricity Trading Arrangements in England and Wales.

Next, section 3.1.3 and 3.1.4 present, respectively, the problem of forecast and control in reinforcement learning theory.

3.1.3 Forecasting

Sutton (1988) presents the temporal-difference learning algorithm as a new method for forecasting. The temporal differences algorithm started-up, back in the 50s, with Samuel's (1959) checkers player, and so far it was in the games research area that the temporal-differences algorithm achieved its highest success: a Gammon player (Tesauro, 1994) and a Chess player (Baxter et al., 2000)[7]. The temporal-difference learning can also be applied to solve control problems (Tesauro, 1992).

Sutton (1988) argues that conventional forecasting methods use the difference between predicted and actual outcomes as their learning signal, while temporal-difference looks at the difference between successive predictions.

Using temporal-difference methods, learning occurs whenever there is a change in predictions. For one-step prediction problems, the best solution is a supervised-learning algorithm. However, for multi-step prediction problems (where a forecast of an outcome at time $t+lag$ is made at time t), the temporal difference methods are best suited, since they learn to correct the predictions (for some *lag*), at time t, taking into account the forecasting errors at time lag-k, for every $k \in [0, lag - t - 1]$. In addition, temporal-difference algorithms can deal with the *credit assignment problem* (to identify, in the set of actions that a player played before a certain outcome, the action which is the responsible for that outcome) more easily than supervised learning can.

Let $V(s_t)$ stand for the value of state s_t, u_{t+1} represent the reward of received after executing action a_t, at state s_t (that resulted in the transition to state s_{t+1}), and Γ denote the trajectory $s_0, s_1, \ldots, s_n$. At the end of the episode, the expected value of

[7] In problems with great state spaces, the computer spends a lot of time doing wasteful exploration (looking at areas of the state space completely irrelevant). Moore and Atkeson (1993) develop a new

each state s_k in the trajectory Γ is updated by the rule[8] $V(s_k) := V(s_k) + \alpha\left[u_{t+1} + \gamma V(s_{k+1}) - V(s_k)\right]$, where *k*, *k+1*, ... represent the index of each state in Γ, and $0 \leq \gamma \leq 1$ and $0 \leq \alpha \leq 1$ denote, respectively, the *discount* parameter and the learning rate.

The temporal-difference algorithm is a *bootstrapping method* as it updates the values of the weights based on previous existing predictions. Equation $d_k \equiv u_{k+1} + \gamma V(s_{k+1}) - V(s_k)$ defines the temporal differences, which represent the difference between the new $\left[u_{k+1} + \gamma V(s_{k+1})\right]$ and the old predicted value of *s*, $V(s_k)$.

The temporal-difference algorithm only updates the values of the states in the trajectory leading to a given outcome, given the values of its successors, and the reward associated with the transition departing from that state. This is a first-reward-based learning algorithm (and thus known as TD(0)), as it uses the value of the next state as a proxy for the value of all the other rewards, i.e., the forecast for the value of each state *k* is based on the one-step Bellman equation $V^{\pi}(s_k) = E\left[u_{t+1} + \gamma V^{\pi}(s_{k+1})\right]$.

The *n*-step Bellman equation $V^{\pi}(s_k) = E\left[\sum_{i=1}^{n}\left(\gamma^{i-1}u_{k+i}\right) + \gamma^{n}V^{\pi}(s_{k+n})\right]$ can approximate the expected value of a state s_k. In this case, the expected value of *n* future rewards and of the value of the state reached after *n* steps is used to forecast the value of a certain state *s* at stage *k* (s_k), under a policy π (Bertsekas and Tsitsiklis, 1996).

heuristic, *prioritised sweeping* that enables the exploration to be concentrated in the parts of the state space that the algorithm expects to be most interesting.

[8] := is an operator used to represent an iterative process. This means that the value of a certain variable is numerically approximated by iterations of the algorithm on the previous value of the same variable.

Moreover, this means that instead of having a one-step updating function it is possible to have a two-step, three-step, or generically an n-step updating function:
$V(s_k) := V(s_k) + \alpha\left[u_{k+1} + \gamma u_{k+2} + \ldots + \gamma^{n-1}u_{k+n} + \gamma^n V(s_{k+n}) - V(s_k)\right]$.

Thus, the difference between a one-step and an n-step TD method is that an agent computes a new estimate of the true value of the state by looking n-steps into the future. If there is no preference on the number of steps, instead of looking several steps ahead, an agent can use instead the weighted average of possible multi-step Bellman equations.

The TD(λ) algorithm in equation (3.3) is a particular way of averaging the multi-step Bellman equations. This average includes all the n-step equations weighted proportionally to λ^{n-1}, where $0 \le \lambda \le 1$.

$$V^\pi(s_k) := V^\pi(s_k) + E\left[\sum_{m=k}^{+\infty} (\gamma\lambda)^{m-k} d_m\right]. \quad (3.3)$$

Equation (3.3) can be approximated by equation (3.4) representing the updating function of the TD(λ) algorithm.

$$V^\pi(s_k) := V^\pi(s_k) + \alpha \sum_{m=k}^{+\infty} (\gamma\lambda)^{m-k} d_m. \quad (3.4)$$

Equations (3.3) and (3.4) are appropriate for off-line problems (in which the value of every state is updated after the trajectory reaches a terminal state), however, they are not adequate for problems where a terminal state does not exist, or in which there are trajectories that may be infinite. In order to apply temporal-differences methods to these types of problems, an on-line TD(λ) has to be defined instead.

In the on-line TD(λ) algorithm there is a memory variable associated with each state, known as *eligible trace*. This variable determines how the process of updating the state-value equation takes into account a certain d_k difference. Eligibility traces are memory parameters, associated with a certain state, that enable the solution of the *temporal credit assignment problem* by giving more credit, i.e., higher trace, to the states visited more recently.

The eligible trace for a state s at time t is denoted as $e(s) \in R^+$:

$$e(s) := \begin{cases} \gamma\lambda e(s) & \text{if } s \neq s_t \\ 1 + \gamma\lambda e(s) & \phantom{\text{if }} s = s_t \end{cases}, \qquad \forall\, s \in S,$$

where s_t is the state of the environment at time t.

The eligibility trace decays for all states at a rate γλ, except for the last state visited where the eligibility trace is incremented by one unit. λ is the *trace decay parameter*, it determines how different states are assigned a certain prediction error, given the discount rate.

Let $d_k \equiv u_{k+1} + \gamma V^\pi(s_{k+1}) - V^\pi(s_k)$ and let $e_k(s)$ represent the eligibility trace of state s at time k. Then, in the on-line TD(λ) algorithm, for every s, equation $V^\pi(s_k) := V^\pi(s_k) + \alpha . e_k(s) . d_k$ approximates the state-value function.

3.1.3. Controlling

In the control problem an agent chooses the action, in a given state, that gives him the optimal, or almost optimal, reward.

The way an agent deals with exploitation classifies his learning behaviour as on-policy or off-policy. Off-policy algorithms may update the value of a given state based on actions other than the actually executed during the episode. During a given trajectory Γ an agent may choose a certain action a', at state s, (because he is exploring), but when updating the estimated value of s, the agent may assume that in future repetitions of the game (also known as episodes) he will choose the action with the higher expected value. On the other hand, on-policy algorithms update the value of a state strictly based on the experience gained from executing some policy. In this case, if in a state s the agent chooses an action a' (because he is exploring) then, when updating the expected value of this state, the agent uses the expected value of action a' (and the value of the state reached by executing it).

The distinction between these two types of policy evaluation has very important implications both on the policies derived and on the convergence properties of the algorithms applied (Singh et al., 2000). The policies derived using on-policy algorithms are *safer* than policies derived using off-policy algorithms. While on-policy algorithms take into account the effects of exploration on the overall reward of the policy, off-policy algorithms take into account only the reward of the potentially optimal policy. Hence, in simulated environments, the off-policy algorithms can find the optimal policy, but their use in real-world problems may have disastrous consequences.

However, if the parameters of a given off-policy algorithm are such that it always follows a greedy policy, and never explores, it may have good performances in real world environments. Nonetheless, in this case, there is no distinction between on-policy and off-policy algorithms, as this distinction relies on the manner the algorithms use exploration.

The convergence properties of off-policy and on-policy algorithms are also different. In an off-policy algorithm, the updating function does not take into account the policy used. Therefore, convergence (e.g., the *Q*-learning presented below) requires the algorithm to try each action (in every state) infinitely. On the contrary, the on-policy algorithm learns the value of the actions actually taken, and converges to the optimum only if, in the limit, it chooses the optimal actions. Hence, in this case, convergence occurs only if the algorithm tries each action (in every state) infinitely often and if, in the limit, the learning algorithm is greedy with probability one.

Q-learning and Sarsa are two different algorithms for on-line control problems. While the former uses on-policy learning, the latter applies off-policy learning. The rest of this section presents these two algorithms in order to illustrate and discuss the application of reinforcement learning to the control problem.

Q-learning is a temporal-differences off-policy algorithm used in the control problem.

Again, let $P_{ss'}^{a}$ stand for the state transition probabilities, representing the probability of the system moving from state s to state s', conditional on the agent choosing action a.

Further, let $Q^*(s,a)$ be an optimal *Q*-factor, i.e., the maximum expected value associated with action *a*, in state *s*, then $Q^*(s,a)=\sum_{s'\in S} P_{ss'}^{a}\left[u_{ss'}^{a}+\gamma V^*(s')\right]$, $\forall s \in S$.

$Q^*(s,a)$ represents the expected value of an action *a*, in a state *s*, assuming that the optimal policy is followed within trajectory Γ. Equation (3.5) represents the

Bellman's equation (actually a set of equations) used to compute the optimal value of each state, given the expected value of every action.

$$V^*(s) = \max_a Q^*(s,a),\ \forall s \in S. \tag{3.5}$$

By combining $Q^*(s,a)$ and $V^*(s)$ the expected value of an action a can be defined, for a state s, as a function of the actions in the following states within Γ, equation (3.6).

$$Q^*(s,a) = \sum_{s' \in S} P_{ss'}^a \left[u_{ss'}^a + \gamma \max_b Q(s',b) \right],\ \forall s \in S. \tag{3.6}$$

In equation (3.6) it is clear that the expected value of an action a, in a state s, assumes that the optimal policy is followed in the future. Watkins and Dayan (1992) show that the one-step Q-learning converges to the optimal action-values with probability one if every action-value is tried infinitely often (in the discrete case).

The value iteration algorithm in equations (3.7) and (3.8) approximates the optimal Q-factors. This updating process follows the assumption that the algorithm, in the future iterations, follows the optimal policy.

$$Q(s,a) := Q(s,a) + \alpha.e(s,a).\left[u_{ss'}^a + \max_b \gamma Q(s',b) - Q(s,a) \right],\ \forall s \in S,\ a \in A(S). \tag{3.7}$$

$$e(s,a):=\begin{cases}\gamma\lambda e(s,a) & \text{if } (s,a)\neq(s_t,a_t)\\ 1+\gamma\lambda e(s,a) & \phantom{\text{if }} (s,a)=(s_t,a_t)\end{cases}. \tag{3.8}$$

An on-policy control algorithm, such as Sarsa, computes the expected value of a policy taking into account the possibility of explorative behaviour (an explorative action may be non-optimal at a certain stage).

In the Sarsa (λ) algorithm, the action value function $Q(s,a)$ is approximated by the updating equation $Q(s,a):=Q(s,a)+\alpha.e(s,a).\left[u^{a}_{ss'}+\gamma Q(s',b)-Q(s,a)\right]$, $\forall s\in S$, $a\in A(S)$, where the eligible trace is computed using equation (3.8).

Therefore, the algorithms of reinforcement learning assume a stationary environment, in which the rewards received from executing a certain action are constant over time. This is quite an inadequate assumption when dealing with multi-agent environments. Typically, several agents learn at the same time and, consequently, adapt their behaviour, i.e., the fact that an agent is learning changes the intrinsic nature of the stochastic processes generating the rewards. Learning generates non-stationary environments, creating the need to model opponents' behaviour.

3.2 Opponent Modelling

The fact that the other agents are also learning modifies the parameters of the game and the value functions that a player is controlling. Thus, the game is non-stationary. Further, not only these parameters are not constant, but also the structure of the game (the transition functions) evolves over time. Therefore, there is a need to look at models where there is common knowledge that the game structure changes and where every agent attempts to learn and adapt simultaneously.

The rest of this section introduces the theory in opponent modelling.

In order for inference of strategic behaviour to be possible, some rules need to regulate the definition of strategies. Rubinstein (1986) proposes the use of finite automata as a tool to model an agent's behaviour. An automaton is a decision rule, or a strategy, consisting of a finite set of states, a transition function (that defines the rules of transition between states) and a behavioural function (defining an agent's behaviour in each state of the automaton). He suggests that repeated games with finite automata could capture a player's bounded rationality (considering automata with a bounded number of states). At the same time, the introduction of finite automata constrains the type of strategies played: only regular strategies are admissible (i.e., given the same input, a player reacts always in the same manner).

Finally, it is noteworthy that, long before Rubinstein proposes the automata game, Schreider (1964) presents the formalism of dynamic programming to solve discrete deterministic problems using finite automata, and introduces its possible application to game theory.

In automata theory there are four major central issues: the complexity of computing the best response automaton, the equilibrium in automata games, the dynamics problem, and finally automata inference.

3.2.1 Best Response Automaton

The complexity of computing the best response automaton aims to capture an agent's rationality and the cost of operating an automaton. To measure this complexity has become a central issue in automata game theory. Rubinstein (1986) uses the number of states in an automaton as a measure of the agent's complexity. Moreover, since the behaviour and the transition functions are costly to operate (this is an assumption), the decision-maker reduces these costs by minimising the number of states in his automaton. Therefore, a player has profit maximisation as his main objective and lexicographically he minimises the number of states in the automaton (i.e., given two automata with the same expected profit the agent chooses the simplest one).

Banks and Sundaram (1990) generalise this measure of complexity in order to include the costs of monitoring the opponent's behaviour. They argue that one needs to measure the "transitional complexity" in order to capture the full complexity of an automaton. Additionally, they compare the complexity of two automata taking into account both the number of states and the number of transitions between them. Moreover, Banks and Sundaram, as well as Piccione (1992), model best response in automata games within a discounted infinite stage model. In this context, Piccione shows the equivalence between the best policy derived by the automaton best response and the stationary policy computed by dynamic programming.

Gilboa (1988) defines the problem faced by every player using a finite automaton: to choose the best-response automaton, given the choices of the other players. He defines Nash equilibrium as the choice of an automaton, for each player, such that no player can increase his payoff by unilaterally changing his automaton. He is particularly interested in the complexity of *computing the best response automaton* rather than in the complexity of implementing it. This is a very strong criticism of the previous literature. He argues that limitations on the number of states do *not* capture bounded rationality, which is a restriction on the capabilities of an agent to *design* a strategy, while the bound on the number of states only captures a limitation on the *implementation* of a strategy. Further, also within this topic, Papadimitriou (1992) shows that there is a trade-off between the designing and the implementation complexities. In other words, the problem of computing the best response automaton with a bound on the possible number of states is *NP-complete* whereas, if no bound is given, this problem can be solved in polynomial time.

Gilboa's work gave rise to a series of papers on the complexity of computing the best response automaton. Ben-Porath (1990) analyses the calculation of the best response automaton under uncertainty (assuming that the opponent has a set of possible automata to use during the game). He shows that, when there is uncertainty concerning the automata the other players use (and he analyses a game against nature), the problem of computing the best response automaton is *NP-Complete*, meaning that there is no polynomial algorithm to solve it. Finally, a probabilistic automaton that chooses actions randomly can also capture uncertainty. Freund et al. (1995) show that, in this case, the complexity of finding the best response automaton against a random opponent is equivalent to the deterministic case.

3.2.2 Equilibrium in Automata Games

The computation of equilibrium is the second major issue addressed by the research in automata games. Abreu and Rubinstein (1988) show that automata games have a Nash equilibrium. This proof is available in a setting where there are two players that play a one-shot game with complete information, and therefore assuming that the players *always use the same automaton*.

Further, Abreu and Rubinstein (1988) show that in a Nash equilibrium of an automata game (with two players), each player uses a finite automaton with an equal number of states. Since the automata are finite, the game eventually reaches a *cycle* where it repeats the pairs of states played. This introduces a partition of state pairs into those belonging to a cycle and those only played at the beginning of the game. The states of a player's automaton that appear in a cycle are all distinct: the other states appear in the beginning of the play are never repeated. In equilibrium, there is a one-to-one correspondence between the stage-game actions of the two automata.

However, even in such a simple setting the use of the Nash equilibrium as a tool of analysis is problematic. Gilboa (1988) criticises the concept of Nash equilibrium as the number of players in a game is usually not known. In addition, Gilboa and Zemel (1989) disapprove of the Nash equilibrium concept in automata games as the problem of determining its existence or its uniqueness is *NP-Hard*.

3.2.3 Automata Dynamics

As recognised by Rubinstein (1998), the automata game, as formulated in this literature, is a one-shot game in which the players cannot alter the automaton and where dynamic aspects are not considered. Therefore, there are no strategic links among the repetitions of the game and a player is not concerned with the impact of his actions on the strategies of the other players.

However, recently, there have been several attempts (coming from the artificial intelligence area) to incorporate dynamic issues into automata games. Carmel and Markovitch (1999) develop an algorithm that enables a player to infer a model of his opponent by interacting with him. This algorithm incorporates an endogenous exploration mechanism that enables a player to plan his actions in advance (an agent modifies his behaviour and learns from the interaction with his opponent). Thus, a player attempts to profit from his knowledge and, at the same time, tries to improve the inferred model when his predictions are incorrect. Nonetheless, this research still assumes a stationary opponent, i.e., it assumes that the opponent does not change his automaton.

Moreover, in a dynamic game, as Mor et al. (1996) put it, a player engages in three tasks at the same time: to define the strategy to play the game, to learn the opponents' strategy, and finally if the other players are also learning, to influence the opponents' beliefs. This line of research assumes that some of the players are not perfectly rational and that their limitations may be learned and exploited by the other players in the game.

3.2.4 Automata Inference

In a dynamic game, a player faces not only the control problem (to optimise his automaton against a given opponent), but also an inference problem (to learn the rule or rules of behaviour used by his opponent or opponents).

There are two major branches in the automata learning literature (Angluin, 1987; Gold, 1978): active and passive learning. A learning algorithm tries to infer the automaton generating a stream of data. Active learning is the inference problem faced by a player who has the ability to influence the input generation process, i.e., he is able to select the inputs for the automaton generating the data. In a passive learning problem, a player has no control over the inputs supplied to an automaton.

Thus, the actions of the learning player, in an active learning algorithm, affect the output of the automaton generating the data. Hence, each player faces an active learning problem. Angluin (1987) shows that a learning algorithm provided with counterexamples and with the possibility of controlling the inputs to the target automaton on any specific type of input, can learn the target automaton in polynomial time.

Furthermore, Angluin also defines the *minimum requirements* to learn the best possible rational model of an automaton's behaviour: *completeness* and *consistency*. The completeness requirement implies that a player holds a forecast for every action in every state of the automaton model. The consistency requirement implies that a player holds a correct model of the automaton he is inferring. A model is correct if, in a certain state, it does not forecast different transitions for the same action. Thus, the completeness and consistency requirements enable a player to infer a model where the transition and behavioural functions are complete and have no contradictions.

However, Angluin's algorithm is not satisfactory to model competitive games as in most game-theoretical models there is a very high cost in experimenting with opponents. This was first realised by Carmel and Markovitch (1996, 1999). They develop an algorithm to infer the automaton used by an opponent. The player inferring his opponent's behaviour uses as only source of information his own experiences in interacting with that given opponent. Carmel and Markovitch's approach deals with an opponent that presents a stationary behaviour.

3.3 Summary

In summary, reinforcement learning can be useful to model problems where an agent plays within a stationary environment, having a high number of opportunities to interact with other players. Furthermore, this technique has the advantage of being able to work in settings where a model of the environment is not available.

However, in real world problems, not only the rewards received from interacting with the environment but also the expectations regarding other player's behaviour are important to define a player's actions. A first shortcoming of reinforcement learning models is the assumption that an agent only learns by interacting with other players. In reality, an agent learns by being told, by observing how others behave, by deduction, and by induction.

Furthermore, a persistent drawback within the reinforcement learning literature is the way it deals with the exploration versus exploitation problem. In all the approaches that have been analysed (n-armed bandit, Q-learning and Sarsa), this is a central issue as it influences the value of the different policies. In addition, if, as discussed, the Q-learning and the Sarsa algorithms present alternative ways of dealing with this problem, all of them still assume that a player is willing and able to try any action he wishes to evaluate, which is clearly a very strong assumption when modelling learning in games.

The second technique presented, automata modelling, enables the design of an agent that infers his opponent's behaviour and exploits this knowledge by creating better strategies to play the game. However, this literature still assumes a stationary opponent, or the existence of an exogenous set of automata from which a player can choose his strategies. Further, it assumes that a player computes the best-response strategy against a given automaton used by his opponent. It is, however, never explained how the players come to share some common knowledge of the set of possible automata.

In Chapter 4, an *n-armed bandit* algorithm models the interaction between agents in a large simulation model of the England and Wales electricity market. Chapter 6, based on the automata games literature, presents the Finite Automata Dynamic Game, which is able to model games having a limited number of interactions between a player and the environment, and games where the trajectory is important for the phenomenon studied.

CHAPTER 4

SIMULATING THE NEW ELECTRICITY TRADING ARRANGEMENTS

The new restructured electricity markets tend to be characterised by an oligopoly of generators, very little demand side elasticity in the short term, and regulatory intervention. Another important characteristic of these markets is the increased uncertainty due to decentralised decision processes such as investment and electricity pricing. These features of this problem suggest that agent-based computational methods could perform a useful role.

Thus, it would seem that the development of a detailed simulation platform representing the agents, the markets, and the market clearing mechanisms, together with reinforcement learning to facilitate profit-seeking behaviour by the agents, could provide a computational framework to overcome the limitations of the analytical approaches.

Furthermore, Roth and Erev (1995) and Erev and Roth (1998) show that the intermediate behaviour of a dynamic game may be more important than its asymptotic properties. They have shown that reinforcement learning outperforms the equilibrium predictions in certain games. In the broader context of management research, the strategy and organisational behaviour literature (e.g., Bettis and Prahalad 1995) has also started to model organisations as adaptive systems that learn with experience, by trial and error, through exploiting and exploring their environments.

However, in practice there have been very few large-scale applications to industry behaviour. It is, therefore, still an open question how well agent-based simulation can provide useful insight at the firm level in a complex market such as electricity, not only with respect to the specification of appropriate learning, but also in terms of analysing the multi-agent and multiple equilibria output from the simulations. This Chapter develops one such application of agent-based simulation to explore the economic implications of the proposed changes to the UK electricity market, and thereby provides a timely case study in using evolutionary computation in practice.

This study was actually motivated early in 2000 by the impending radical reforms of the electricity market (which were scheduled to be introduced in November 2000). The New Electricity Trading Arrangements (NETA) would replace the mandatory, daily uniform price auction (the "Pool"), which had operated to provide a competitive wholesale market since 1990. The new market is mainly based upon continuous bilateral trading up to "gate closure" at 3.5 hours ahead of real-time. Following gate closure, the system operator would operate a balancing mechanism, the design of which was deliberately intended to financially reward flexible plant and discourage players, on both the generation and demand-side of the market, from being out of balance. The risks involved in this new system, therefore, appeared to be considerable and any model-based insights would have to capture subtle details of the inter-relationship of the bilateral trading to the balancing market, and the relative plant economics that would follow. With no realistic analogies from electricity markets elsewhere, and only some limited simulations from an experimental role-playing game, specifically from London Economics (1999, October), agent-based simulation appeared to offer a real possibility to develop detailed insights into the potential market ahead of its introduction (which actually happened in March 2001).

Previously, Bower and Bunn (2000, 2001) applied agent-based simulation to look at one aspect of the auction design involved in NETA, namely the switch from uniform to discriminatory pricing. Bower and Bunn's agent-based platform enables a detailed description of the market, taking into account discrete supply functions, different marginal costs for each technology, and the interactions between different generators. However, it does not capture the interaction between the bilateral trading and the balancing market, nor does it incorporate any sophistication in the agents' learning abilities. The simulation platform developed here is a much more detailed representation of how NETA was designed to work:

1. It actively models the demand side (suppliers).
2. It models the interactions between two different markets (the bilateral market and the balancing mechanism) and the settlement process.
3. It takes into accounts the daily dynamic constraints, and it assumes different marginal costs for each generation technology.

4.1. Specification of the NETA Simulation Platform

4.1.1 Overview of NETA

The conceptualised agents in this model represent generators (i.e., generation companies possibly owning several plants with different generation technologies), suppliers (i.e., agents purchasing from the wholesale market in order to "supply" end-use customers) and the system operator (SO). Next, NETA is described together with each one of its components: the bilateral trading, the Balancing Mechanism (BM), the Settlement Process (SP), and the strategic behaviour of suppliers and generators. The basic structure of the model follows Ofgem (1999, July), Ofgem/DTI (1999, October) and London Economics (1999, October), although the details of the BM reflect later revisions Ofgem (2000, March) and Ofgem (2000, April). Table 4.1 summarises the NETA structure.

TABLE 4.1: The NETA Structure

1. Forward and Futures Market • Futures and forward contracts • Years, months, days ahead • Generators and suppliers buy and/or sell electricity • Future and forward prices	**3. Balancing Mechanism** • 3.5 hours ahead • The SO accepts offers or bids from generators and suppliers • Pay as bid auction
2. Day-Ahead Power Exchanges • One day-ahead trading • Generators and suppliers buy and/or sell electricity • Power exchange prices	**4. Settlement Process** • The SO charges the companies if contract positions ≠ metered volumes • System sell price • System buy price

Table 4.2 presents the day ahead, and within day, balancing schedules (OFGEM, 2000, April: 160), which are summarised in Figure 4.1.

TABLE 4.2: Day ahead and within day balancing schedules

The SO day ahead balancing process:

1. *By 09:00 publishes the day ahead demand forecast.*
2. *By 11:00 receives the Initial Physical Notifications (IPNs).*
3. *Calculates the available national plant margin or shortfall.*
4. *Verifies system security given demand predictions, the submitted IPNs and planned transmission outage.*
5. *By 12:00 issues the total system plant margin data to the market for the day ahead.*
6. *Forecasts constraint costs based on the estimated Final Physical Notifications (FPNs) and bid (offer) prices and volumes.*
7. *If necessary calls the most economic Balancing Services contracts to ensure the system security.*
8. *During the following eleven hours, receives updates of the Physical Notifications (PNs).*
9. *By 16:00 publishes the revised national plant margin and zonal margin.*

The SO within day balancing process:

1. *Publishes half hourly averaged demand forecasts for a defined period, until gate closure.*
2. *As participants become aware of changes to their physical position, they advise the SO.*
3. *At defined times the zonal and national margins are reassessed and provided to the market.*
4. *Undertakes security analysis and reassess the requirements of Balancing Services contracts.*
5. *At gate closure, the PNs become FPNs and the SO will have received bids (offers) of the prices and volumes from the participants in the BM.*
6. *During the BM, the SO balances the system taking into account the technical constraints, the dynamic operating characteristics of generation and the demand balancing services.*

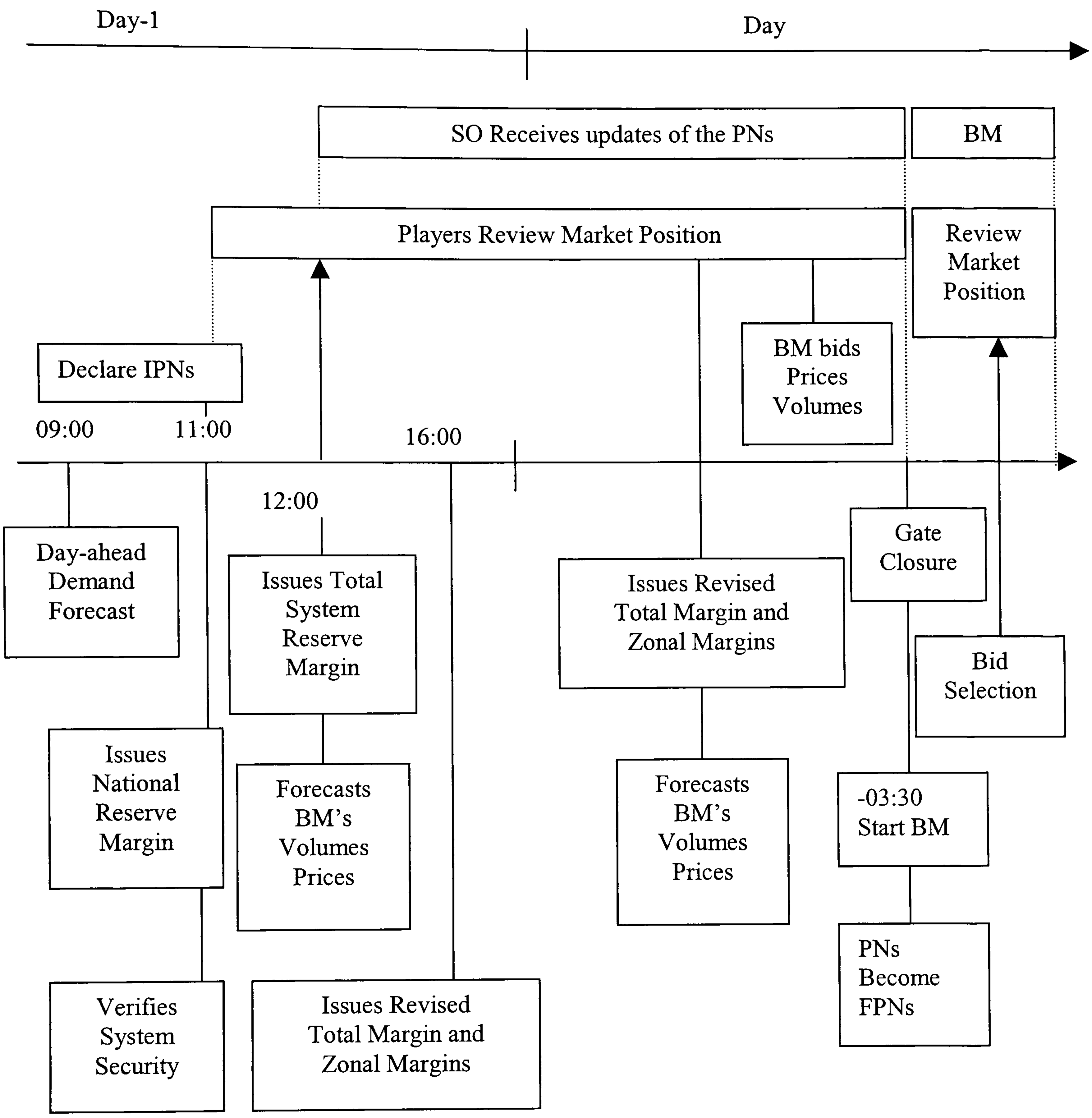


FIGURE 4.1: The SO balancing process

NETA is based upon unadministered bilateral trading between generators, suppliers, traders and customers, which takes place in the forward markets before gate closure. At the introduction of NETA, four organisations had set up power exchanges to facilitate this, although it is clear that market forces will cause liquidity to gravitate to one or two of them. The BM works as a market where the SO buys and sells increments (*incs*) or decrements (*decs*) of electricity in order to balance the system as a whole. However, individual generators and suppliers may be out of balance. During the SP, to calculate the *Imbalances*, the SO compares, for each one of the suppliers and generators (plant by plant), the contract positions (quantities contracted) plus whatever each plant bought or sold in the BM, with the actual position (quantities generated or consumed). The Imbalance may be a *spillage* (if a plant is generating more than it has contracted, or if a supplier is consuming less than it has contracted) or a *top-up* (if a plant is generating less than it has contracted, or if a supplier is consuming more than it has contracted). For both types of imbalances, there is a price: if a plant is spilling, it receives as payment for the electricity generated the System Sell Price (*SSP*); if a plant is topping-up, it pays the System Buy Price (*SBP*). The spread between the two prices intends to provide a penalty for being out of balance. The regulatory authority and the government expect the *SSP* (*SBP*) to be considerably lower (higher) than the prices in the forward markets.

The variables used in this model are defined as follows.

QD_t: total quantity of demand at time t;

QPX_t: total quantity of demand at time t in the bilateral markets;

QTP_t: total quantity of top-up at time t;

$QBBM_t$: total quantity bought in the BM at time t;

QG_t: total quantity of generation at time t;

QS_t: total quantity of spillage at time t;

$QSBM_t$: total quantity sold in the BM at time t.

Then, the demand (equation (4.1)) and generation (equation (4.2)) identities describe the interactions between the two markets:

$$QD_t = QPX_t + QTP_t + QBBM_t \tag{4.1}$$

$$QG_t = QPX_t + QS_t + QSBM_t. \quad (4.2)$$

The SO's task is to buy or sell in the BM enough electricity to ensure that the system is always in balance, $QSBM_t - QBBM_t = QTP_t - QS_t$. During the BM, the SO balances the system taking into account the technical constraints, the dynamic operating characteristics of generation and demand balancing services, and the demand uncertainty.

The SO starts the balancing process one day ahead of the trading time by publishing the demand forecasts, both for the system as a whole and for each zone in the network. The SO revises these forecasts after taking into account both the weather and the network conditions, and the Physical Notifications by generators and suppliers. The SO also forecasts constraint costs, based on the estimated *Final Physical Notifications* (*FPNs*) and bid (offer) prices, and volumes, in the BM. If necessary, the SO calls the economic Balancing Services contracts to ensure the system security.

Within the day balancing process, the SO revises his demand forecasts every half-hour, for each defined period, until gate closure. The participants may revise their physical notifications until gate closure. At gate closure, the *PNs* become *FPNs* and the SO receives bids (offers) of the prices and volumes from the participants in the BM. During the BM, the SO balances the system, taking into account the technical constraints, the dynamic operating characteristics of generation and demand balancing services, and the uncertainty in demand.

4.1.2 Modelling NETA

The first concern in the development of a simulation model aimed at market design and policy analysis is to determine if the conceptual model it is an *accurate* representation of the system studied, i.e., to determine its validity (Law and Kelton, 1991: 298).

The validation of the NETA platform is an important and difficult task as the main objective of this work is the analysis of market design. Specifically the analysis of the new market rules that specifically aimed to change the way the E&W Pool was working. Thus the previous system could not be used as a source to test the model.

Therefore, this model was validated by:

1. Following the descriptions of how the future market was going to work and the discussions within the industry, and trying to replicate the "mental model" to which the regulator seemed to be converging at the time. Hence, the basic structure of the model follows Ofgem (1999, July) and Ofgem/DTI (1999, October), although the details of the BM reflect also the later revisions Ofgem (2000, March) and Ofgem (2000, April).

2. Seeking help from experts in the industry. This model was developed with the active participation of several experts from one company in the E&W electricity market, involving consultations and direct advice from traders, analysts and directors. Further, in the design of this model were also actively involved some of the experts employed by the regulator in the development of NETA.

3. Applying existing theory to model the players' behaviour. The algorithms used to capture the players' behaviour were grounded on economic learning

theory, and on the common sense of the experts in the industry regarding the management of plant portfolios:

a. Generators' behaviour was design to model the way these players could use a learning algorithm to manage plant portfolios, to learn the trade-off between risk and return, and to learn the interactions between the balancing mechanism and the bilateral markets.

b. Supplier's behaviour was design to capture the way a company would manage risk, and learn the interaction between bilateral markets and balancing mechanism.

4. Collecting the parameters related to the fundamentals of the industry from reliable sources. This parameters respect the marginal costs, retail price and the technological characteristics of each genset in the industry, given the fuel prices in 2000.

5. Finally, the prices in the several experiments run with this model, most of which are not presented in this thesis, were consistent with the experts expectations of how the NETA would work in the future.

However, building a detailed model of the balancing process and its interactions with the power exchange and futures market would present overwhelming complexities and so, for tractability, some simplifications were adopted (which are much less restrictive than the ones usually presented in the literature):

1. These simulations do not model the transmission system. This implies that the model does not capture regional imbalances or transmission constraints.

2. The model simulates only a typical day, taking into account some plant dynamics. The main goal of this model is to analyse the process of finding the equilibrium solution for the specific daily profile simulated.

3. The continuous nature of trading in NETA is simplified. The forward market and the BM are represented as two sequential one-shot markets. This implies that the flows of information are much simpler than reality: the players only submit their offer (bids) to the bilateral market knowing the SO*'s* demand forecasts, and submit their offers (bids) to the BM knowing the expected position of the system.

4. Finally, the model assumes independence between generators and suppliers, although vertical integration is a reality in the industry. This simplification is adopted as the regulatory authority imposes a condition that all the trading between the suppliers and generators belonging to the same company has to be subject to separate imbalance accounts.

In formulating the market models, the "natural" approach would seem to be the Continuous Double Auction (Milgrom, 1987; Wilson, 1987), however its implementation with computer agents has several drawbacks. It requires direct communication and negotiation between agents; it demands the development of a multi-criteria algorithm when the quantity is variable; and in a large model such as the one presented in this chapter, computational time is an issue[9].

Hence, instead, the NETA model uses the Single Call Market developed by Cason and Friedman (1997). However, it is adapted to reflect the NETA trading principles, with the agents paying the price bid or receiving the price offered, instead of paying (receiving) the clearing price. The option for pay-as-bid reflects the intuition that under bilateral trading the agent's need to learn how to offer the electricity of each plant, e.g., a nuclear plant cannot bid zero or a negative price, as it used to happen in the Pool.

It is noteworthy that von der Fehr and Harbord (1993) use the same type of auction, which they call a sealed-bid multiple-unit auction. Nicolaisen, Petrov and Tesfatsion (2001) adopt an auction very similar to the auction used in the NETA model, with the main difference being that in their model an agent pays (receives) the midpoint of the bid-ask spread.

However, the option for the Single Call Market assumes equilibrium trading, and therefore it does not allow the analysis of some possibly important implications of moving from a pool to a bilateral market. Weber (1995), in the context of modelling a stock market, shows that moving from a centralised to a decentralised market mechanism can imply increased trading costs and more importantly increased market opacity (which may drive some trades very far from equilibrium). This is clearly a concern in liberalised electricity markets (specially in the bilateral market case). However, for efficiency reasons, and as this is not the research question of this chapter this effect is not analysed here.

Another challenging task is the modelling of suppliers and generators as bounded rational agents. In this case, each player should be able to learn from the environment and improve his behaviour by interacting with it, thereby avoiding the perfect foresight paradigm. However, at the same time, following Marcet and Nicolini (2002, December)[10], an agent's behaviour should not be too non-rational (some lower bounds on the agent's level of rationality were imposed). The internal consistency requirement is the most demanding one. An agent does not choose completely unreasonable actions, even in the early stages of the learning process.

Thus, the specification of the computational agents needs to take into account how they manage information. An agent takes actions in two different markets (the power exchange and the BM) and then he receives feedback from these two markets. In order to facilitate the association between an action and its results, the feedback

[9] In this section, each experiment simulates 200 trading days. In each trading day, there are 24 auctions for each hour in the bilateral market, and 24 auctions for the BM (one for each hour).

[10] They define three concepts used to establish lower bounds on rationality. In asymptotic rationality, the agents' behaviour will asymptotically converge to the optimum. In epsilon-delta rationality, agents may have some resistance to change, and may be satisfied within a solution close enough to the optimal behaviour. With internal consistency, an agent's behaviour has to have some lower rational bounds in the short-run, i.e., an agent tries to do the best he can over a limited horizon.

process should be as close to the action as possible. In order to accommodate this requirement an agent has *operational objectives*, taking into account, however, that his final goals are the *strategic objectives*. Finally, *operational rules* are defined to impose the minimum rationality requirements, and to relate *strategic* and *operational objectives*.

The model is organised into trading days, or iterations (the NETA algorithm is presented in Table 4.3). A trading day starts with the PX trade. In this market, every supplier tries to buy, at a price as low as possible, the amount of electricity needed to fulfil his contract cover objective (the amount of electricity he wants to buy in the PX). On the other hand, every generator tries to sell at a price as high as possible, given his portfolio of plants, and the amount of electricity he wants to save to sell in the BM. After the trading in the PX ends, each agent knows exactly how much he has sold (or bought) and he knows the *FPNs* of each one of his plants, in each one of the hours. Overall, in each day (iteration) 48 auctions take place. Per hour, there is an auction in the PX and another one in the BM.

TABLE 4.3: NETA Algorithm

For the number of trading days specified

1. Suppliers predict demand for each hour

2. Generators define which plants can run

3. Generators offer in the PX

4. Suppliers bid in the PX

5. Trading in the PX and calculation of System Position in each one of the hours

6. Generators and Suppliers offer (bid) into the BM

7. Trading in the BM and calculation of imbalance prices

8. Settlement Process: calculation of imbalances for each one of the suppliers and generators (plant by plant)

9. Learning

Then, the trading in the BM begins. Each supplier, given his expectations, sells or buys in the BM, in order to avoid imbalances. Each generator knows the amount contracted for each one of his plants and the system position, long or short, and decides if he sells or buys in the BM. It is noteworthy that given a certain "system position" only a certain type of offers (increments or decrements) is accepted. The trades in the BM occur between the SO and each one of the generators and suppliers offering (bidding) in the BM. The SO accepts bids or offers from suppliers only for arbitrage reasons, otherwise he takes the offers or bids from generators. After all the trading in the BM is finished, the final positions and imbalances (of each generator and supplier) are calculated. Finally, the SO computes the imbalance prices and costs.

At the end of each trading day, each agent tries to learn from the experiences he has accumulated in the previous days by evaluating the profit received in the PX and in the BM. Then he defines *new policies* to bid or offer in the PX and in the BM, in order to maximise his expected profit. Each agent has some general *strategic objectives* that he tries to achieve, namely profit maximisation and exposure to the BM. An agent learns to improve his performance by using reinforcement learning (for a survey in reinforcement learning see Kaelbling et al., 1996; Sutton and Barto, 1998).

In this formulation, the following notation is used:

C_{ij} represents the short-run cost of plant j, owned by a generator i.

PO_{ij}, $PObm_{ij}$, ($PBbm_{ij}$) represent, respectively, the price offered in the PX and the price offered (bid) in the BM, by generator i for plant j.

Q_{ij}, Qx_{ij} and $QObm_{ij}$ ($QBbm_{ij}$) represent, respectively, the total electricity sold by plant j, the quantity sold in the PX and the quantity sold (bought) in the BM.

PB_i ($PBbm_i$) represents the price bid in the PX (BM) by supplier *i*.

S_i, Sx_i, Sbm_i ($SObm_i$), Imb_i represent, respectively, the market share, the quantity bought in the PX, the quantity bought (sold) in the BM, and the imbalance of supplier *i*.

D and P_{retail}, $PBbm_i$ ($PObm_i$) represent, respectively, the total demand at a certain time, the retailing price and the price bid (offered) by a supplier buying (selling) in the BM.

The objective of each player is to maximise his (average daily) total profit, taking into account the assets he owns, the total demand, the behaviour of other players, and his objective for the BM exposure (BMX_i). What is the intuition behind the BM exposure constraint? The significant gap between the prices in the BM and in the forward and power exchange markets acts as an incentive for suppliers to avoid the BM (the *SBP* (*SSP*) is significantly higher (lower) than the clearing price in the bilateral markets). A generator is less constrained by the BM system. If there is a shortage of electricity, he may sell in the BM (receiving the price offered, if accepted by the SO). If there is excessive supply of electricity, the generator may also bid to reduce his generation (being paid the price bid, if accepted). So, why would a generator prefer to sell in the PX? First, a player may prefer to sell the generation of a baseload plant before gate closure, due to his risk aversion against any interruption in continuous running. Second, since the suppliers have an incentive to trade in the forward and power exchange markets, the shoulder technologies face a high demand in these markets. In contrast, it seems that the BM is the *natural environment* for expensive and flexible peak-plants looking for price spikes. The problems faced by generators and suppliers are separately analysed in section 4.1.2.a) and 4.1.2.b).

4.1.2.a) The Suppliers

A supplier tries to find the best possible policy (which price to choose), given the behaviour of the other players. Formally, a supplier *i* tries to maximise the total daily profit, equations (4.3), by learning iteratively the prices to bid (or offer) in the PX and in the BM, and at the same time respecting his objective for the BM. Table 4.4 describes the iterative process by which a supplier tries to approximate the equilibrium represented by equations (4.3).

$$Max\ \pi_i = \begin{cases} \mathrm{Sx_i}.(\mathrm{P_{retail}} - PB_i) + \mathrm{Sbm_i}.(\mathrm{P_{retail}} - PBbm_i) + \\ +SO b\mathrm{m_i}.PObm_i + \mathrm{Imb}_i.(\mathrm{P_{retail}} - SBP) & if \quad \mathrm{Imb_i} \geq 0 \\ \\ \mathrm{Sx_i}.(\mathrm{P_{retail}} - PB_i) + \mathrm{Sbm_i}.(\mathrm{P_{retail}} - PBbm_i) + \\ +SO b\mathrm{m_i}.PObm_i - \mathrm{Imb}_i.SSP & if \quad \mathrm{Imb_i} < 0 \end{cases} \quad (4.3)$$

st.

$$\mathrm{S_i} = \mathrm{Sx_i} + \mathrm{Sbm_i} - SOb\mathrm{m_i} + \mathrm{Imb}_i.$$

$$\mathrm{Sbm_i} + SOb\mathrm{m_i} + |\mathrm{Imb}_i| \leq BMX_i$$

TABLE 4.4: Supplier's learning algorithm

1. Verifies that the objective for the BM exposure was achieved

If this is not the case, penalise the profit obtained in the PX.

2. Given the daily profits in the PX and in the BM revises the:

2.a) Expected profit in the PX and in the BM

2.b) Expected acceptance rate in the PX and in the BM

3. Defines the new bidding policy for the next day.

4. In the beginning of the next day, after calculating predictions for demand, defines the quantities and prices bid.

A player uses reinforcement learning to compute the expected profit and acceptance rate in Table 4.4 and Table 4.5. In this algorithm, the learning is through the *accumulated* history of previous plays. At each one of the iterations, a player updates the value of the expected profits and acceptance rates of each possible mark-up, and decides between following the best policy so far and exploring an alternative one.

Each supplier is characterised by the following pre-specified parameters: *Market share, Balancing Mechanism Exposure (contract cover), Retail Price, Prediction Error,* and *Search Propensity.*

Market Share refers to the relative quantity of electricity sold in the retail market. *Balancing Mechanism Exposure* reflects the percentage of forecast demand a supplier intends to purchase in the PX. *Retail price* is not a crucial variable, since the goal of this model is short-term analysis. However, it allows the modelling of a supplier's problem as profit maximisation. *Prediction Error* reflects, as mean average prediction error, the capability of the agent to predict his own demand. *Search Propensity* is an integer ranging from one to ten, which defines a heuristic control of how the agents search for the best payoff and transform past experience into future policies. A low *Search Propensity* generates stable policies that change slowly with experience, whereas a high *Search Propensity* generates reactive policies that tend to follow short run experience).

Each supplier has the following instruments:

1. *Mark-up* in the PX (learned by the agent). This is relative to the *PX price* (*PXP*) in the previous day.

2. *Mark-up* in the BM (learned by the agent). This is relative to the *PXP*, in the same day.

Each supplier has two *strategic objectives*:

1. To maximise *total daily profits*, given the market structure and its market share.

2. To minimise the difference between his *objective* for the BM exposure and the *actual* BM exposure.

The two *operational objectives* are to maximise daily profits in the PX and in the BM, respectively. They enable each agent to associate the outcomes (such as prices, quantities, and profits) to the instrument and *strategic objectives*.

Suppliers also have an operational rule based upon adaptive expectations: never bid (offer) more (less) in the BM than the previous day's *SBP* (*SSP*) as those are the expected imbalance prices[11].

4.1.2.b) The Generators

A generator faces the problem of finding the best possible policy (how to price his plants), given the behaviour of competitors and suppliers. Formally, a generator i tries to maximise daily profit (π_i), equations (4.4), by iteratively learning the prices to offer (or bid) in the PX and in the BM (as a simplification, in this representation it is assumed that a generator is never imbalanced).

[11] This operational rule, also adopted for the generators, establishes lower bounds on rationality. One of the agents' strategic objectives is BM exposure. Agents want to have a low BM exposure because of high BM price and volume uncertainty. If suppliers (generators) choose to pay (receive) a higher (lower) price in the PX than in the BM their behaviour would be irrational since this behaviour would contradict their BM exposure objective.

$$Max\ \pi_i = \sum_j \mathrm{Qx}_{ij}.\left(PO_{ij} - C_{ij}\right) + \mathrm{QObm}_{ij}.\left(PObm_{ij} - C_{ij}\right) - \mathrm{QBbm}_{ij}.\left(PBbm_{ij} - C_{ij}\right)$$

$$\begin{aligned} & st. \\ & \mathrm{Q}_{ij} = \mathrm{Qx}_{ij} + \mathrm{QObm}_{ij} - \mathrm{QBbm}_{ij} \\ & \sum_j \mathrm{QObm}_{ij} + \mathrm{QBbm}_{ij} \leq BMX_i \end{aligned} \tag{4.4}$$

TABLE 4.5: Generator's learning algorithm

For each plant in the portfolio:

1. Verifies the objectives for the BM exposure and if they were not achieved, penalises the profit obtained.

2. Given the daily profits in the PX and in the BM, revises for the mark-ups used, the:

2.a) Expected profit in the PX and in the BM.

2.b) Expected acceptance rate in the PX and in the BM.

3. Defines the new offering policy for the next day.

4. In the beginning of the next day, defines the quantities and prices offered for each plant after knowing which plants are available, making sure that all the operational constraints are respected.

Table 4.5 describes the iterative process by which a generator tries to approximate the equilibrium represented by equations (4.4).

Each generator is characterised by the following parameters: *Plants, Cycles, Capacity, Availability, Balancing Mechanism Exposure (contract cover)* and *Search Propensity*.

Plants stands for the generation units owned by a generator. By definition, plants of the same type have similar marginal costs, start-up costs and no-load costs. The agent's decision-making does not take into account explicitly the start-up and no-load costs although he implicitly learns to do so. Thus, each agent has an objective for the position of each plant in the load duration curve and his profits are penalised whenever a plant does not meet this objective.

This is a reasonable way of incorporating some consideration of dynamic plant constraints. Therefore, the parameter *Cycles* is defined for each type of plant, see Table 4.6. This classification considers: (a) the peak technologies (gas turbine, oil and pumped storage) to be able to run a maximum of three times a day; (b) the shoulder technologies to run a maximum of two cycles (they include CCGT and coal); (c) the base-load plants to run in a non-stop regime (these are the nuclear stations and the interconnectors).

TABLE 4.6: Relation between Cycles and type of plant in the UK

Type	Cycle
Gas Turbine	3
Oil	3
Pumped Storage	3
Small coal	2
CCGT	1
Large coal	1
Inter-connector	0
Nuclear	0

Thus, baseload plants with high start-ups or inflexible technology run continuously or with just one cycle. Flexible plant, which have low start-up cost, can exhibit a higher number of cycles. *Availability* defines the probability of the installed *Capacity*, i.e., it reflects outage rates for each plant. *Balancing Mechanism Exposure* and *Search Propensity* are the same as the ones defined for suppliers.

Each generator has the same instruments as the suppliers. Specifically, each generator learns to mark-up in the PX from the previous day and to mark-up in the BM from the PX outcome, and his *strategic objectives* are:

1. To maximise *daily profits*, given the market structure.
2. To minimise the difference between his *objective* and the *actual* BM exposure, for each plant.

The two *operational objectives* are to maximise the *daily profits* in the PX and in the BM, respectively, for each plant.

In order to avoid inconsistent behaviour during the learning process, the model imposes some lower bounds on rationality through *operational rules*:

1. *Portfolio Management:* A plant with higher or equal number of cycles never undercuts offers of another plant with equal or less number of cycles.
2. *No interruption*: Plants that have to run continuously, or plants with one cycle, may run without profit in certain hours of the day.
3. *No Loss-leading*: Plants with one cycle do not run without profit at the beginning or at the end of the day. They prefer not to run at all if the price is to low.
4. *Peak Premia*: Peak plants never offer prices below marginal cost.

Together with two extra operational rules for the BM:

1. *Adaptive Expectation:* Never bid (offer) above (below) the previous *SBP* (*SSP*).
2. *Avoidable Cost*: Never pay more than the marginal cost for "speculative" *decs*.

4.1.3 Simulating Balancing Mechanisms

At gate closure, each agent knows exactly how much he has sold, or bought, and provides the *SO* with his *Final Physical Notification* (*FPN*)[12]. Furthermore, the SO*'s* total demand forecasts are assumed common knowledge in the industry at every iteration (and in these experiments these forecasts are also assumed to be accurate). Nevertheless, each supplier has some uncertainty when forecasting his own demand.

Then, using the *FPNs* and his demand forecast, the SO calculates the total system surplus, or shortfall, for each period in the day ahead. Given this total system position the SO accepts either increments (*incs*) or decrements (*decs*) in the BM. The trades in the BM occur between the SO and each generator and supplier offering (bidding) the *incs* or *decs* into the BM. Further, in the interest of efficiency the SO clears all possible *arbitrage opportunities* (ARB_t). Then he decides if he should accept extra bids or offers by computing the difference between contracted demand and supply ($Excess_t$):

$$Excess_t = QD_t - FPNs_t - ARB_t - QS_t + QTP_t.$$

If $Excess_t > 0(< 0)$ the SO accepts *extra-incs* (*extra-decs*) above the arbitrage level ARB_t from generators (only[13]).

After all trading in the BM is finished, and the SO bought or sold the quantity of electricity needed to balance the system, he computes the imbalances of each generator and supplier, and the imbalance prices and costs. If the SO accepts *incs,* the *SBP* is defined as the weighted-average of the offers accepted in the BM. Otherwise,

---

[12] *Good behaviour assumption*: this model assumes that agents always communicate the true *FPNs* and that they always deliver what the quantities contracted in the PX and in the BM.

[13] Note that a supplier without load management will not influence the net position of the system, and so his bids in the BM are only to avoid the imbalance charges. Only if there is an arbitrage opportunity are these bids (offers) accepted.

if the SO accepts *decs,* the *SSP* is defined as the weighted-average of the bids accepted in the BM. Hence, in our model, at each hour, only one imbalance price is defined (*SBP* or *SSP*). Thus, if an agent is long (short) when the system is short (long) there is no imbalance price. In this case, the algorithm assumes that if there are insufficient bids (offers)[14] the *SO* takes the average of past *SBP* (*SSP*) values for that particular hour.

4.2. The Learning Algorithm

Although there are several agent-based concepts coming from the field of multi-agent systems (e.g., Franklin and Graesser, 1997), almost all assume that an agent has communication capabilities and that he is a physical entity (i.e., a computer process). The type of agent used in the NETA model has no communication capabilities. This agent is a conceptual identity representing an "economic agent" in the market, and having the ability to "receive" information, learn from the interactions and act on the simulated environment. This is an *autonomous agent*[15].

Moreover, each agent in the NETA model learns on-line (given the information set, he modifies his actions in order to maximise his own profit). Thus, a natural way to model the type of learning in a market such as NETA seems to be reinforcement learning (see Chapter 3).

This algorithm differs from Roth and Erev's (1995) algorithm as:

[14] In practice, this will not occur as frequently as in the model. The dynamic nature of buying and selling by the SO throughout the 3.5-hour BM window, and the locational nature of buying and selling to balance the system at the nodal level, means that the SO will be more active than our model suggests.

[15] An autonomous agent is a system situated within a part of an environment that senses that environment and acts on it, over time, in pursuit of his own agenda and to affect what he senses in the future (Franklin and Graesser, 1997).

1. *It is a pure strategy stochastic game.* An agent tries to learn the best pure strategy, and the learning process is organised in order to find pure strategy equilibria

2. *It takes into account the exploration vs. exploitation problem.* In Erev-Roth's model, these concepts are not present. An agent using reinforcement learning defines how much exploration and exploitation to undertake. In the model defined in Table 4.7, an agent only employs, with a high probability, the best strategy, or the first three best strategies, and only with a low probability does he apply the other ones. The learning parameter α is also very important, as it is equal to 0.5. This implies that this models a tracking problem (an agent is able to quickly react and change strategy).

3. *A player always has the same probability of exploitation or exploration independently of the iteration of the game.* This approach keeps the agent's capacity of reaction to changes constant within a simulation, therefore making the model suitable to deal with non-stationary environments. In Roth-Erev's basic model the probability of choosing a certain strategy is directly proportional to the expected reward. This implies that Roth-Erev's algorithm tends to converge to local equilibria with insufficient search of the solution space.

4.2.1 Defining the mark-ups

Let $up_t(mk,a)$ denote the mark-up bid or offered, at time *t*, in market *mk*, which equals PX or BM and denotes the power exchange and the balancing mechanism, and where *a* equals *B* or *O* denoting a bid or an offer.

A mark-up in the PX denotes the ratio between the price bid (offered) in the PX and the *PXP* in the previous day. A mark-up in the BM denotes the ratio between the price bid (offered) in the BM and the *PXP* in the same day (the agent already has information on the *PXP* when he bids or offers in the BM).

The attraction of this model is its simplicity: an agent only learns a ratio for the whole day, and the mark-up revision in the PX or BM is the same for each different hour of the day. This avoids the adoption of different learning behaviour for the different levels of demand (von der Fehr and Harbold, 1993).

Another strong point about this learning algorithm is the assumption that the information set includes only the prices in the previous day. Given the present state of the world an agent tries to improve his position given the knowledge accumulated in the past (only), and therefore assuming that the present contains all the relevant information (i.e., the Markov property).

Let t represent an iteration (K is the maximum number of iterations), $t=1,\dots,K$, and i represent the time of the day, $i=1,\dots, 24$. Let also PB_{ti} $(PBbm_{ti})$ represent the price bid, at iteration t and hour i, in the PX (BM), and PO_{ti} $(PObm_{ti})$ represent the price offered, at iteration t and hour i, in the PX (BM). Then:

$$
\begin{aligned}
PB_{t,i} &= PXP_{t-1,i}.up_t(PX,B) \\
PO_{t,i} &= PXP_{t-1,i}.up_t(PX,O) \\
PBbm_{t,i} &= PXP_{t-1,i}.up_t(BM,B) \\
PObm_{t,i} &= PXP_{t-1,i}.up_t(BM,O)
\end{aligned}
.
$$

Each agent learns a different policy for each one of the four mark-ups. This is a very important point about this learning algorithm. An agent does not learn how to choose the prices; instead, he learns how to choose the mark-ups on the previous day's prices. The result is a model where prices are unbounded.

The learning process is the same for each one of the four mark-ups (which are partitioned into discrete intervals). In the experiments described in this chapter, the mark-ups assume different ranges for suppliers and generators in different markets.

A supplier learns different mark-ups for the PX and the BM (one for *incs* and another for *decs*)[16]:

1. Bids in the PX from 0.95 to 1.2: this allows for decreases and increases in the price offered.
2. Incs in the BM from -0.2 to 3: this allows for incs with negative prices.
3. Decs in the BM from -0.2 to 3: this allows for decs with negative prices.

A generator also learns different mark-ups for the PX and the BM (one for incs and another one for decs):

1. Offers in the PX from -0.15 to 1.15: this allows for decreases and increases in the price offered. It also allows the generators to buy in the PX (by offering a negative price).
2. Incs in the BM from 0.6 to 3: this allows incs to be higher or lower than the PX price.
3. Decs in the BM from -0.25 to 1.1: this allows for decs with negative prices.

The larger range defined for a generator's mark-ups allows them to bid above or below the clearing price, and allows some plants to receive a negative price in order to make sure some low cycle plant runs through short price troughs.

[16] The intervals are constructed in a fairly ad-hoc way. They were defined in order to allow enough freedom in the possible choices (allowing for increases and decreases in offers and bids, or infinitely negative or infinitely positive bids or offers).

4.2.2 Learning a Policy

At each instance, an agent computes the *expected daily profit* and the *expected acceptance rate* for each one of the mark-ups used at that specific iteration:

1. The expected daily profit is the exponential smoothing of the profits earned in the past trading days.
2. The expected acceptance rate is the exponential smoothing of the number of successful bids (offers) per day.

Thus, given the expected acceptance rate and the expected daily profit, each player calculates the *expected reward* for each mark-up. Subsequently the agent constructs a *utility function* over these mark-ups.

In order to construct this utility function the agent ranks the mark-ups by decreasing value of expected reward. The mark-up with the highest expected reward receives a higher perceived utility value. This transformation also takes into account the S*earch Propensity* parameter such that a low parameter is associated with a conservative utility function. A high search parameter, in contrast, defines a more adventurous agent, that is able to try different mark-ups. Then, the agent transforms the utility function into a *policy*, which is an association between each mark-up and the probability of bidding (offering) that mark-up. This agent uses the policy to choose the bid (offer) price in the following day. The basic ideas behind the construction of the utility function and the derivation of the policy are inspired by the fitness function and selection mechanisms which have been used in genetic algorithms (Michalewicz, 1992; Whitley and Kauth, 1988).

Let j=1,..., 10 represent the interval index (the mark-up number) and t = 1..., K denote the iteration, and define the following variables.

Prf_{tj} (Art_{tj}): daily profit (acceptance rate) at iteration *t*, using mark-up *j*;

$E(\text{Prf}_{tj})$: expected daily profit for time *t* (conditional on acceptance) of mark-up *j*[17].

$E(\text{Art}_{tj})$: expected time *t* acceptance rate, using mark-up *j*.

$E(\text{Rwd}_{tj})$: expected time *t* reward of mark-up *j*.

Rank(j) : stands for the rank of mark-up *j*.

Util_j and Pol_j : perceived utility and the probability of using a mark-up *j*.

The derivation of the policy follows the algorithm in Table 4.7.

At the end of the day, after receiving the feedback with the prices and quantities traded in each hour, each player computes the policy to use in the next trading day. First, a player calculates the new expected daily profit and acceptance rate for the mark-ups used (which he ranks by decreasing value of expected reward).

TABLE 4.7: Learning a Policy

After each iteration:

1. *Compute the expected daily profit for each mark-up j used at time t*

$$\text{Prf}_{tj} = \sum_{i=1}^{24} \text{Profit}^i_{tj}$$

$$E(\text{Prf}_{tj}) = E(\text{Prf}_{t-1j}) + \alpha.[\text{Prf}_{t\text{-}1j} - E(\text{Prf}_{t-1j})]$$

2. *Compute the daily acceptance rate for each used mark-up j at time t*

$$\text{Art}_{tj} = \frac{\text{Number of bids (offers) accepted}_{tj}}{24}.$$

$$E(\text{Art}_{tj}) = E(\text{Art}_{t-1j}) + \alpha.[\text{Art}_{t\text{-}1j} - E(\text{Art}_{t-1j})]$$

[17] In this notation $E(X_{tj})$ stands for the expected time t value of the variable X, at time t-1.

3. *For each used mark-up j recalculate the expected reward at time t*

$$E(\mathrm{Rwd}_{tj}) = E(\mathrm{Prf}_{tj}).E(\mathrm{Art}_{tj})$$

4. *Rank the mark-ups by decreasing value of expected reward*
5. *Compute the perceived utility of each mark-up j*

$$\mathrm{Util}_j = U \cdot \left(\frac{\mathit{Search\ Propensity} - d}{\mathit{Search\ Propensity}} \right)^{\mathit{Rank(j)}-1}$$

6. *Compute the policy*

$$\mathrm{Pol}_j = \frac{\mathrm{Util}_j}{\sum_k \mathrm{Util}_k}$$

After computing the expected reward, a player calculates the utility received from each one of the mark-ups. This utility, in this specific model, is not directly dependent on the rewards but on an agent's search propensity and on the specific rank of each mark-up.

In the experiments presented in section 4.3, for each agent, the parameters *U*, *Search Propensity* and *d* are equal to 1000, 4 and 3, respectively.

This approach is quite flexible, enabling the construction of a wide variety of utility functions. This learning algorithm is a central point in the NETA model. By avoiding to follow the standard *n-armed bandit* algorithm (section 3.1.2), but instead following the approach in genetic algorithms that aims to avoid fast convergence to local minima, the NETA learning algorithm avoids one of the problems in reinforcement learning: the inability to deal with non-stationary environments. The NETA learning algorithm, by not directly relying on the actual rewards received but depending on the mark-up rankings instead, is able to model systems that have non-stationary rewards.

After calculating the perceived utility from each mark-up, the agent transforms this utility function into a *policy*, i.e., the probability of using each mark-up *j*. For this purpose, the algorithm applies the rule of proportionality. The probability of choosing a certain mark-up is directly proportional to the weight of the perceived utility of that mark-up to the sum of perceived utilities of all mark-ups.

The example in Table 4.8 illustrates the learning process. An agent holds ten different possible mark-ups and computes a policy on them. $E(Prf_{tj})$, $E(Rwd_{tj})$, and $E(Art_{tj})$ represent, respectively, the expected time *t* profit, and the expected time *t* reward, and the expected time *t* acceptance rate, of each mark-up *j*. *Rank(j)* orders the rewards from the highest to the lowest expected value. $Util_j$ represents the perceived utility an agent receives from a certain mark-up. Pol_j represents the probability of using a certain mark-up when bidding (offering).

TABLE 4.8: Policy derivation example

Mark-up Categories	1	2	3	4	5	6	7	8	9	10
$E(Prf_{tj})$	500	400	600	300	1000	700	800	850	750	900
$E(Art_{tj})$ (%)	100	94	98	80	85	70	65	70	60	55
$E(Rwd_{tj})$	500	376	588	240	850	490	520	595	450	495
Rank(j)	5	9	3	10	1	7	4	2	8	6
$Util_j$	3.9	0	62.5	0	1000	0.3	15.6	250	0.1	1
Pol_j (%)	0.3	0	4.7	0	75	0	1.2	18.8	0	0.1

4.3. Initial Simulations of the New Electricity Trading Arrangements

The NETA simulation model, as described in the previous sections, was applied to the full system of E&W electricity industry as it existed in summer 2000, with 80 generating plants, owned by 24 generators who sell power to 13 suppliers. All the experiments have simulated 200 iterations (trading days) of the algorithm, based upon the winter demand profile shown in Figure 4.2, with an available capacity of around 56 GW.

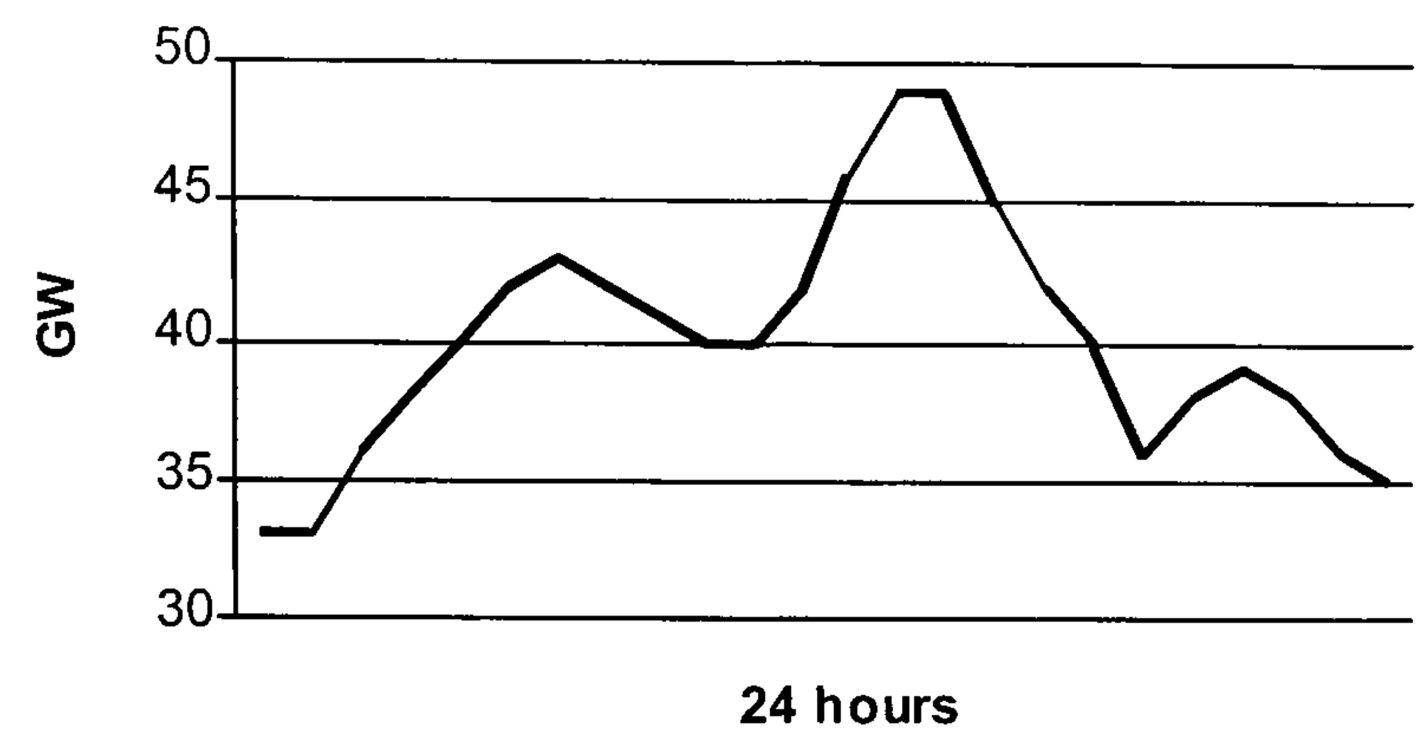


FIGURE 4.2: Standard winter daily demand profile

The demand profile used in the experiments presented in this chapter is characterised by presenting three peaks of demand. In these experiments, the learning rate is 0.5 and the retail price is £80. Further, plant availability is a function of the specific technology. Thus, baseload technology is available 99% of the days, while the flexible technology is available 95% of the days.

The estimated *marginal generation costs* for each plant ranged from £3 MWh to £94 MWh[18]. The low marginal costs are associated with baseload plants, nuclear, CCGT and some large coal plants. The high marginal costs are associated with gas turbines, pumped storage and oil plants. The estimates used are consistent with those used in other published studies on the UK generation market, as well as with known data on plant efficiencies and fuel costs.

Table 4.9 presents the market shares of the eight largest generators in the E&W electricity market (in 2000).

[18] The generation database contains capacities and marginal costs for every genset (see Table A.1).

TABLE 4.9: Generation Industry Structure

Capacity of each Company (% of Total, 59 GW) in 2000				
	Total	Nuclear	Large Coal+ CCGT	Small Coal +OCGT + OIL + Pump. Storage
PG	16.5		19.7	24.9
NP	13.9		16.3	22.5
BE	12.4	54.0	4.9	
Edison	10.6		10.1	30.7
TXU	9.7		11.6	14.7
AES	7.8		10.1	6.8
EDF	4.7	17.3	2.0	
Magnox	3.9	19.9		
Others	20.5	8.8	25.3	0.4
Total GW	59.1	11.4	40.7	7.0

The source for Table 4.9 is the UK Electricity Association. The first source is the Electricity Association's (2000a) brief history of the UK electricity sector. The second source is the Electricity Association's (2000b) description of all the plants in the UK. The third source is the Electricity Association's (2000c) list of company ownership. The final source is the (Electricity Association's (1999) analysis of the industry's evolution during the 90's.

Table 4.9 splits the generation capacity owned by each generator into three categories by taking into account the degree of flexibility and running times of each technology. The first category includes nuclear plants, which run continuously. The second category includes large coal and CCGT (the shoulder technologies). Finally, the third group includes small coal, OCGT, oil and pumped storage (the peak plants). The capability of a generator to exercise market power depends on his market share and on the technologies owned. It is important to look at the relative market share of each generator in each type of technology in order to understand its *strategic position*. Thus, BE has 54% of the nuclear generation capacity installed in E&W (and 4.9% of the shoulder capacity), while AES owns both shoulder (10.1% of shoulder capacity) and peak plant (6.8% of peak capacity).

Table 4.10 depicts the market shares for the retail supply business (in 2000).

TABLE 4.10: Supply Industry Structure

Peak Demand 50 GW in 2000 Market Share of each Company (%)			
London Electricity	12.2	PG	10.2
Eastern	11.2	Scot. & Southern	9.9
NP	10.2	Others	40.7

The supply business was less concentrated than the generation business. The biggest supplier was London Electricity with 12% of market share. This was the input used in the simulations presented in this chapter. In reality, vertical integration between generation and supply, and local oligopoly games, makes this analysis more complex than the one captured by the model.

Figure 4.3 depicts the relation between *cycles* and a plant's position in the supply curve, i.e., it plots the type of cycle defined for each plant within the supply curve. On average, flexible plants have higher marginal costs.

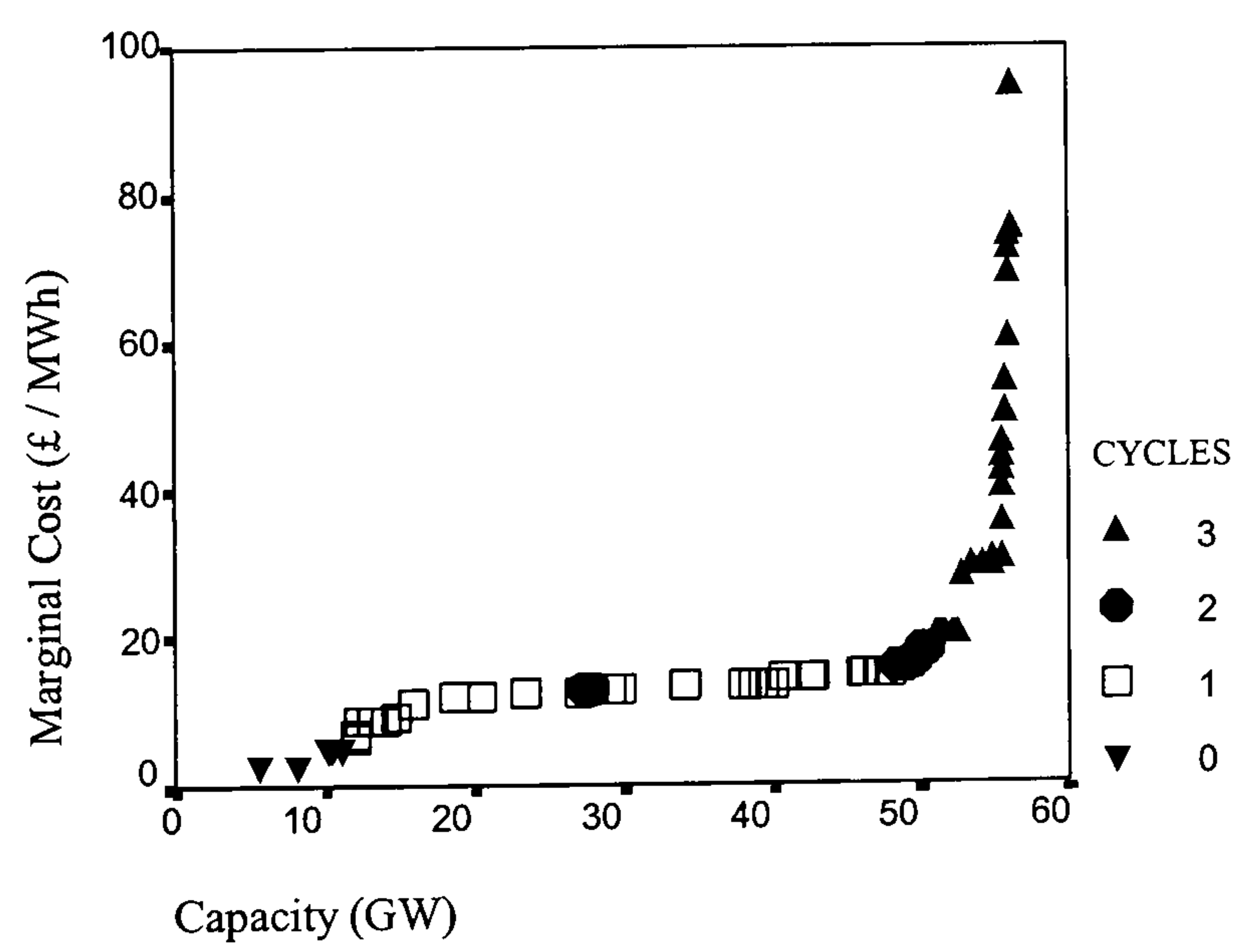


FIGURE 4.3: Plant cycles in the supply curve.

4.3.1 Market Results

The initial results give an overview of how the model works and the evolution of prices and quantities traded in the PX and BM. In Table 4.11, *Sell* (*Buy*) BM stands for the quantity sold (bought) in the BM. *Spill* (*Top-up*) is the total electricity that suppliers consumed below (above) the *FPNs* or that generators produced above (below) the *FPNs*. This table presents all quantities as a percentage of the total trade in the PX.

TABLE 4.11: Quantities traded in the BM and total *Imbalances*

Average Prediction Error	Sell BM	Buy BM	Spill	Top-up
10.0	2.4	1.3	5.1	6.2
5.0	1.3	0.9	2.5	2.9

The quantity traded (bought or sold) in the BM is less than 2.5% of the total trade in the PX. As one might expect, the prediction error is a determinant factor in the amount traded, and these results point to the economic gain that the suppliers could achieve through better forecasting (they will be paying the imbalance charges resulting from prediction errors). Whilst forecasting load at the national level has become very accurate (2% mean absolute percentage error), at the regional level, where metering is less frequent and where in the past there had not been a need to forecast accurately in such a short time-scale, errors of between 5% and 10% are thought to be the norm. Imbalances and trade in the BM are different concepts. The spillage and top-up represent from 2.5% to 2.9% of total trade in the PX with a low prediction error, and represent 5.0% to 6.1% of total trade in the PX for the higher prediction error. The reason why imbalances represent a higher percentage than trades in the BM is that some positive imbalances can cancel the negative ones. Overall, these results show that the risks of NETA are much greater for suppliers.

The generators can control their imbalances, with the exception of unplanned technical outages, while the suppliers are completely dependent on industry prediction capabilities.

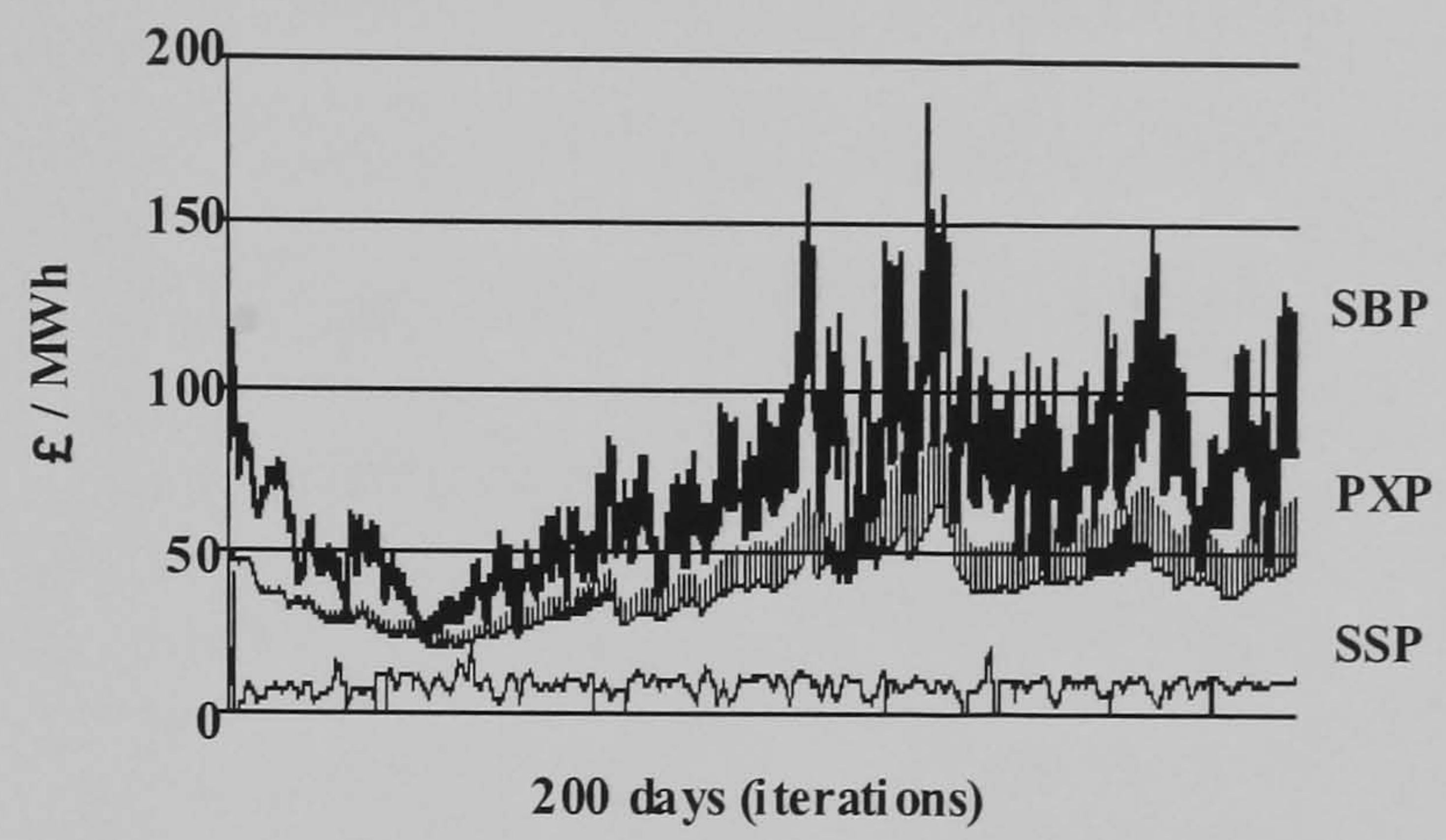


FIGURE 4.4: Average daily prices (moving 24 hours window).

Figure 4.4 displays, as an illustration, a representation of the average daily prices (24 hour moving window) in the PX and in the BM (*PXP*, *SBP*, and *SSP*), during the 200 iterations of the baseline experiment. Note that market prices do emerge to create a wide spread between *SBP* and *SSP*, and that *PXP* is centrally located between them. This is what the advocates of NETA hoped to occur: out-of-balance players would regret they had not traded forward at *PXP*.

To look at the daily price profile in more detail, Table 4.12 shows the actual prices, describing the quantity demanded, the marginal costs and prices (Power Exchange Price and the System Buy Price) during the last iteration, for each hour of the day.

TABLE 4.12: Final Iteration for the Winter Day Baseline Scenario

Demand (GW)	Marginal Cost (MC)	Capacity Used (%)	*PXP*	*SBP*	Mark-up *PXP*/MC (%)	Ratio *SBP*/*PXP*	Mark-up *SBP*/MC (%)
33	13.25	58.8	14.55	16.57	9.8	1.14	25
33	13.25	58.8	14.53	18.00	9.7	1.24	36
36	13.28	64.2	14.89	17.79	12.2	1.19	34
38	13.36	67.7	15.44	19.00	15.6	1.23	42
40	14.32	71.3	15.71	20.00	9.7	1.27	40
42	14.36	74.8	16.99	34.32	18.3	2.02	139
43	14.60	76.6	17.72	34.16	21.4	1.93	134
42	14.36	74.8	16.88	30.85	17.6	1.83	115
41	14.36	73.1	16.03	29.27	11.6	1.83	104
40	14.32	71.3	15.72	20.00	9.8	1.27	40
40	14.32	71.3	15.78	26.50	10.2	1.68	85
42	14.36	74.8	16.71	33.28	16.4	1.99	132
46	14.62	82.0	29.46	63.00	101.5	2.14	331
49	15.87	87.3	318.72	603.70	1908.0	1.89	3703
49	15.87	87.3	509.18	1100.76	3108.0	2.15	6830
45	14.42	80.2	23.65	51.49	64.1	2.18	257
42	14.36	74.8	16.29	20.00	13.5	1.23	39
40	14.32	71.3	15.67	26.60	9.4	1.70	86
36	13.28	64.2	15.15	16.62	14.1	1.10	25
38	13.36	67.7	15.47	18.55	15.8	1.20	39
39	13.36	69.5	15.60	26.80	16.8	1.72	101
38	13.36	67.7	15.49	24.78	16.0	1.60	86
36	13.28	64.2	14.96	19.00	12.7	1.27	43
35	13.28	62.4	14.83	18.00	11.7	1.21	36

The level of prices which have emerged around £50/MWh for winter days is rather high, because in this experiment agents learn to exercise market power and create price spikes at the two peak periods. The mark-ups represent the percent increase of the *PXP* and of the *SBP* compared to the marginal cost. Notice that the *SBP* is much more volatile than the *PXP* or the *SSP*, and the emergence of daily price cycles (with high prices at the peaks) as agents learn from experience.

The peak prices spikes are noteworthy, as the *PXP* ranges from £14.5 to £29.5 in the off-peak hours to £318 and £509 in the peak hours. The mark-ups $\left(\frac{PXP}{MC}\right)$ above

60% are associated with *capacity used* above 80%, and the price explosion only happens with *capacity used* above 85%.

Therefore, this winter trading day presents three distinct periods:

1. A low demand period (below 80% of *available capacity*) where the prices tend to be close to marginal costs.

2. An average demand period (between 80% and 85% of *available capacity*) where the prices are at least 60% above the marginal costs.

3. A high demand period (above 85% of *available capacity*) where the prices rise to at least 38 times the marginal cost.

It is noteworthy that the small difference in *capacity used* between the high demand and the average demand period is due to the shape of the supply function.

These results are stronger than the ones in von der Fehr and Harbord (1993). The experiments show that even if the demand is lower than the total capacity, prices may rise well above the marginal cost. Thus, there is a threshold above which collusion is possible and sustainable, and another above which only the regulator can stop prices from unbound spiking.

The rest of this analysis disregards these two hours with demand of 49 GW, since in these periods the game did not converge to equilibrium. It is also interesting to note that the off-peak prices (without these two hours) show a profile of *PXPs* consistent[19] with the 6-month forward prices, which were emerging in summer 2000, ahead of the planned NETA introduction during the winter (and indeed close to what did occur in the early months of NETA after March 2001).

[19] In August 2000, forward prices for off-peak winter 2000/2001 were on average at £22/MWh.

4.3.b. Strategic Implications for Suppliers

The 13 suppliers included in the experiments had market shares between 3.6% and 12.2%. This section analyses two main issues: demand prediction capabilities and contract cover (how much of their expected demand they want to buy in the PX). Table 4.13 represents the combination of parameters used in these experiments.

TABLE 4.13: Experiments with suppliers' parameters

Exp. Number	Prediction Error	Contract Cover
1	10	100
2	10	115
3	5	115
4	5	100
5	5	85

The influence of the separate effects is now tested. This analysis proceeds by looking at the off-peak hours. Table 4.14 reports the results of a regression where the dependent variable is a supplier's *Profit-per-unit-sold*. In this regression, *Two*, *Three*, *Four* and *Five* are dummy variables representing experiments 2, 3, 4 and 5, respectively (i.e., *Two* is '1' to indicate experiment two or '0' otherwise). The first experiment, *One,* is the base case. Table 4.14 presents the coefficients with the *t*-statistics in brackets.

TABLE 4.14: Dependent variable: suppliers profit per-unit sold

		Two	Three	Four	Five
Coef.	60.8	-5.79	-11.87	0.71	2.4
t-stat	(396)	(-26.7)	(-54.8)	(3.2)	(11.2)

The regression in Table 4.14 shows that the lower the prediction error, the higher the *Profit-per-unit-sold* of the industry as a whole, given an objective for contract cover of 100%. The experiments have also shown that a contract cover of 115% (85%) has a negative (positive) impact on the suppliers *Profit-per-unit-sold*. This result surprised the supply industry since conventional wisdom suggested that a supplier would be wise to be risk averse to imbalances and if anything would be slightly over-contracted in his *FPNs*.

In this model, however, the intuition is that if all suppliers under-contract, the PX price falls given the reduction in demand. Following that, the generators expect a lower price and extra capacity available in the BM, and so the BM price falls as well.

4.3.3 Strategic Implications for Generation

The analysis in this section uses two dependent variables, the *profit-per-unit-of-capacity-available* and price offered in the PX (*price-offered-PX*), to study how generators behave. The objective is to identify a relation between a generator's behaviour and plant ownership structure.

While Table 4.15 (Model 1) presents the parameters of a regression for the *profits-per-unit-of-capacity-available*, Table 4.15 (Model 2) shows the parameters for *price-offered-PX*[20]. In both models, the variables are representative of the accepted offers in the off-peak hours.

The independent variables are dummy variables for the companies, e.g., AES equals '1' if AES owns that plant and '0' otherwise. Further, C1, C2 and C3 are also dummy variables that identify plants with one, two, and three cycles, respectively. Table 4.15 presents the coefficients with the *t*-statistics in brackets.

[20] Although some of the estimated regression parameters are not statistically significant, the overall results do provide an indication for the informal clustering into strategic groups.

TABLE 4.15: The dummy variables AES, BE, EP, MG, NP, PG and TXU identify the seven main owners of plant portfolios

Var.	Model 1	Model 2
	5.9	16.7
	(57.2)	(168)
C1	-2.8	0.3
	(-27.6)	(2.9)
C2	0.16	1.3
	(0.87)	(7.5)
C3	4.3	4.8
	(14)	(16.8)
AES	-1.4	-0.7
	(-12.2)	(-6.4)
BE	1.23	-0.3
	(9.8)	(-2.6)
EDF	1.9	-0.4
	(15.5)	(-3.3)
EP	-0.05	-0.3
	(-0.34)	(-2.2)
MG	-0.4	-0.2
	(-2.1)	(-1.4)
NP	-0.15	0.9
	(-0.8)	(5.6)
PG	-2.1	-0.3
	(-17.2)	(-2.9)
TXU	0.5	2.6
	(2.4)	(12.5)

Next, this analysis proceeds by clustering the parameters of Model 1 and Model 2, and by classifying the generators into strategic groups. Specifically, four different categories emerged.

1. The first category includes AES, MG and PG. These generators have lower profits per-unit and lower prices than average. These players own baseload plants that they use intensively with low profits per unit of capacity available.

2. The second category includes BE and EDF. These generators have a low price and policy of continuous running plants.

3. The third category includes EP. This generator has a portfolio based on baseload plants (large coals) with "high" marginal costs and a few very flexible and "low" cost plants (pumped storage). EP has a very good performance on pumped storage and a below average performance on large coals.

4. The fourth category includes NP and TXU. These generators belong to the same structural group, characterised by "diverse portfolios with dominant positions". These agents have a policy of high pricing (prices above average).

The flexible technology (two and three cycles) tends to exhibit price and profits per unit above average. *The fact that a plant belongs to a certain portfolio of plant, owned by different generators, also has a significant effect on the way its owner prices generation.* This suggests that the inter-relatedness of plant ownership and profit continues to promote the active buying and selling of plants through the capital markets, as generators seek to reposition their portfolios of plants. This plant trading process is the motivation of Chapter 7 (the plant trading game) and, more indirectly of Chapter 6 (the Finite Automata Dynamic Game).

4.4. Equilibrium Analysis

In continuous markets such as NETA, where capacity constraints are a reality, multiple transient equilibria are expect to appear, and thus it would be expected that an evolutionary model might display non-convergent behaviour under some conditions. One way to analyse the equilibrium properties of the NETA model is to repeat an experiment N times and check if the prices converge. A second way is to run the experiments for a longer number of iterations.

To gain insight into the equilibrium properties of this model, FIGURES 4.5 and 4.6 present an extensive analysis of the basic scenario (a winter day with peak demand of 49 GW, and total capacity of 59 GW). These simulations repeat the same scenario 10 times with 400 trading days each. FIGURES 4.5 and 4.6 present the average, maximum, minimum, and the difference between maximum and minimum prices.

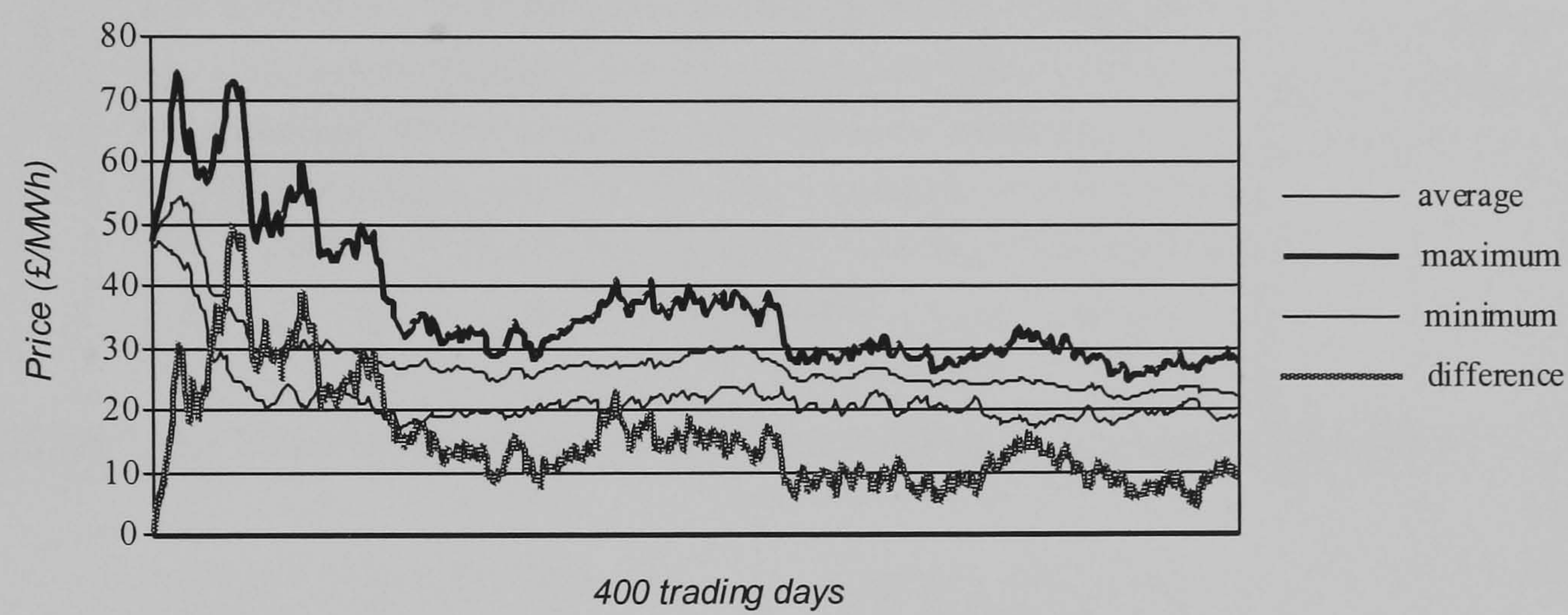


FIGURE 4.5: Price behaviour in an hour with high demand (45 GW)

The analysis of Figure 4.5 suggests that the average prices converge to equilibrium. For example, with 50 days, the prices are in the range of £18.6/MWh and £34.1MWh, with an average of £24.6/MWh. With 400 trading days the prices range from £17.6 to £22.7/MWh, with an average of £19.8/MWh. As the gap between the maximum and the minimum is decreasing, this does suggest convergence.

The analysis of Figure 4.6 suggests that, for the peak demand hours, the process presents two distinct periods. In the first period, prices tend to increase steadily. In the second period, the average price stabilises and the difference between maximum and minimum starts to have a more unstable behaviour with increasing variance, and no convergence (even in simulations with more than 400 trading days). This is not really a surprise. Many theoretical analyses have pointed to the problem of deriving an equilibrium solution for the peak. In the absence of sufficient demand elasticity, the focal point for strategic behaviour becomes the result of a repeated game and longer-term considerations of regulatory and new-entrance consequences.

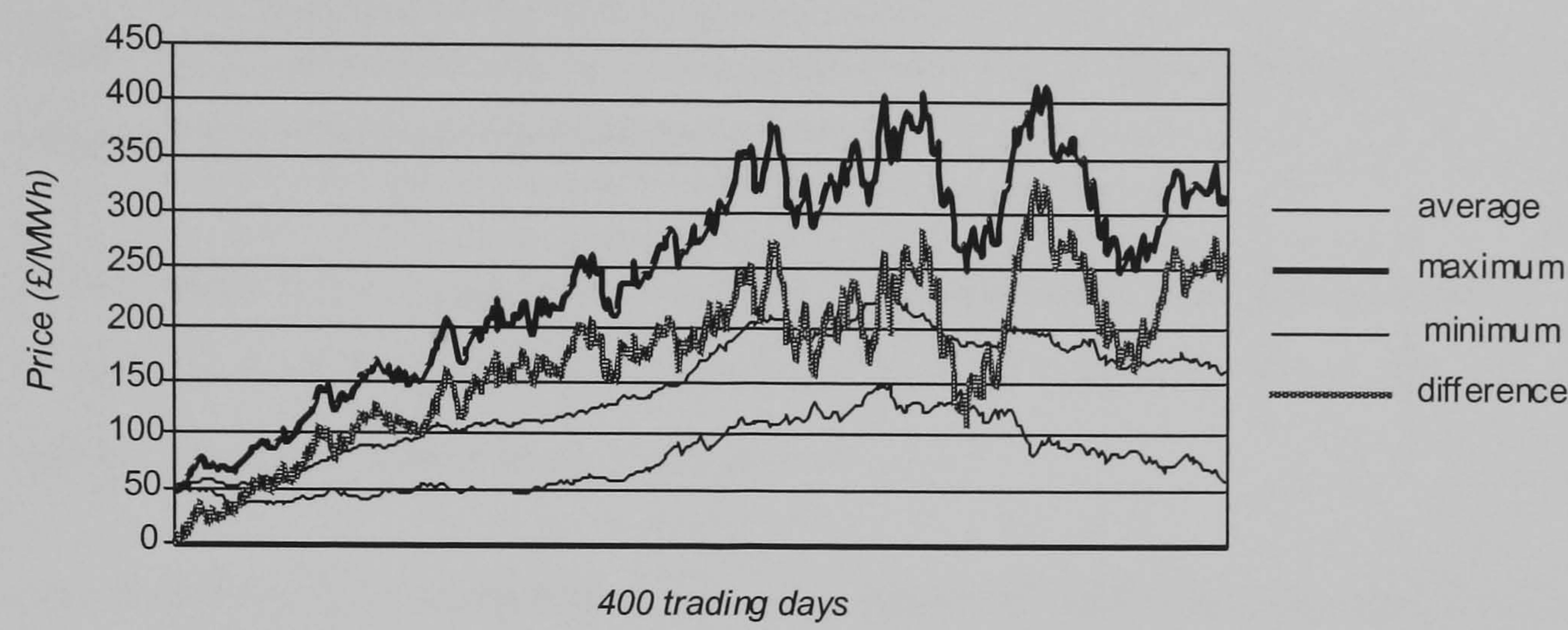


FIGURE 4.6: Price behaviour in the hour with peak demand (49 GW)

4.5. Conclusions

Agent-based computational methods, even when a simple algorithm is used, can provide insights into the pricing and strategic behaviour of electricity companies in the liberalised electricity markets. Imperfect competition, exercised through the daily repetition of a competitive market with administered market rules, creates a process of continuous experimentation and gaming which agent-based simulation is able to mimic.

Overall, the results obtained in this study were plausible to the industry. Further, the model represented the first detailed study of the relation between the bilateral trading and the balancing mechanism in the proposed New Electricity Trading Arrangements for England and Wales.

In summary, the insights into NETA, which came out of this modelling, were:

1. The amount of trading in the BM may be less than 2% of the total amount sold in the bilateral markets. The imbalances, however, will be much higher (the magnitude depends on the suppliers' prediction errors).

2. There is no incentive for the SO to publish better predictions, yet this would improve the efficiency of the industry.

3. NETA will have a higher risk impact on suppliers than on generators, even though one of the goals of the reform was to make the wholesale electricity market less of a "generators' market".

4. The Capacity Margin (demand as a percentage of total capacity available) seems to be the most important factor behind the possibility of collusive behaviour on the generation side. There is a threshold above which collusion is possible and sustainable, and another above which only the regulator can restrain unbounded spikes. Other experiments, not reported in this chapter, on the parameter *availability*, have shown that withholding capacity may have an extreme impact on the peak prices (nevertheless, in the experiments reported, a tightening of the capacity margin by 5% had an important impact on prices).

5. The experiments have also shown that a contract cover of 115% (85%) has a negative (positive) effect on the suppliers' *profit-per-unit-sold*. If they collectively learn to under-contract in this manner, they may exert some market power on the supply side.

6. Flexible plant will be relatively more valuable in the new market mechanism, yet its value will depend upon the portfolio of ownership. This relation will ensure that there will be continued activity in buying and selling plant amongst generators as they seek to reposition themselves in the market.

Next, Chapter 5 addresses the issue of unilateral market power by simulating the behaviour of generators in the NETA platform.

CHAPTER 5

EVALUATING INDIVIDUAL MARKET POWER IN ELECTRICITY MARKETS

With the widespread restructuring of electricity industries, moderating the abuse of market power has become a persistent issue for the various regulatory agencies appointed to manage these liberalised markets. For a number of plausible reasons, including preserving security of supply, and (or) enhancing privatisation returns, the restructuring of electricity markets has resulted, in most cases, in too few players to ensure market efficiency. Therefore, regulators have often inherited the *structural* deficiency of imperfect competition. Under such circumstances, with one or more generating companies evidently possessing market power, monitoring *conduct* and setting a framework for preventing the abuse of market power is a major and on-going methodological task.

While the main concern in Chapter 4 is the analysis and understanding of the issue of market design, and its implications on the companies' strategic behaviour, the issue analysed on this topic is the one of market power, specifically focusing on the ability of an individual company to abuse of their strategic position in an electricity market.

The specific motivation for the study reported here resulted from an initiative by the regulatory agency in Britain (Ofgem) to require all "significant" generators in the market to sign a market abuse licence condition (MALC), which was essentially a "good behaviour clause" in their licence to generate electricity. This had followed a decade of attempts during the 1990s to reduce market power structurally by enforcing divestiture of assets, with relatively little effect (at least by 1999) on wholesale prices, and was being sought (in 2000) ahead of the proposed New Electricity Trading Arrangements (NETA), which was introduced in 2001. Two generators had refused to sign the MALC on the basis that, in their argument, they could not exert a significant influence on market prices. One, British Energy (BE), was quite a large generator with over 10% of market share, but they being almost all baseload nuclear claimed that they rarely set market prices. The other, AES, had below 10% of market share. Because of their refusal to sign the document, the case was referred to the Competition Commission for analysis. This specific modelling challenge was therefore to assess, ex-ante, if such players could really influence market prices profitably by attempting strategic withholding behaviour. Ofgem were clearly anxious that NETA would be introduced with strong restrictions to preclude high prices.

This chapter applies the NETA simulation platform to analyse if the two companies in the Competition Commission Inquiry had enough market power to operate against the public interest. This research specifically analyses the capability of these companies to increase market prices. The Competition Commission defined market power as the ability of a generator, acting independently, to raise prices consistently and profitably above competitive levels.

It is important to take into account some of the important characteristics of electricity markets in general and of NETA in particular, in order to answer the specific question regarding these companies' conduct. Thus, it would seem that the NETA simulation platform would have the level of detail needed for this study.

However, although computational approaches have the capability of allowing the analysis of detailed market models with multiple agents that learn various strategic parameters, they can also pose a major problem in understanding and explaining the outcomes of the evolutionary simulations.

As a first contribution, this chapter presents a stylised Bertrand game with constraints, which explores the main strategic decisions faced by generators in the market and motivates the issue of conduct-evaluation. The aim of this stylised model is to clarify the exposition of the main argument, and to show why, in order to understand more fully the processes governing a market's behaviour, one needs to use some carefully formulated computational simulations. This stylised Bertrand game adapts Crampes and Creti's (2001, October) framework to a pay-as-bid auction with N-generators.

Second, this chapter illustrates the basic capability of agent-based simulation to give insights and a competitive baseline for market power analysis. In addition, the study addresses the specific Competition Commission's questions of whether the apparently small company (AES) and (or) the baseload generator (BE) have enough market power to operate against the public interest.

5.1 Conduct-evaluation, Equilibria and Learning

This section concentrates on the issue of conduct-evaluation, analysing if a generator may find it profitable to price and withhold capacity strategically, without directly addressing the market design relation to competitive behaviour.

This study extends Crampes and Creti's (2001, October) one-shot Bertrand game with capacity constraints to a pay-as-bid *N*-players setting (each player pays the price bid or receives the price offered).

This model assumes that: a) There is only one market, the power exchange (disregarding the interactions with the BM). b) Each generator only owns one type of technology, characterised by a certain short-run production cost, which is assumed constant. c) Demand is inelastic, *D*. d) There are *N* generators *a*, *b*, …, *N*, with short-run generation costs C_a, C_b, …, C_N, and capacities K_a, K_b, …, K_N. e) *D* is such that $\sum_{i=a}^{N} K_i > D$ and if K_s is the capacity of the smallest generator, $\sum_{i=a}^{N} K_i < D + K_s$. f) A generator, in the maximum, can charge a price $\overline{P}$.

The Bertrand game with capacity constraints is a two-stage model. In the first stage a generator chooses the amount of capacity available and, in the second stage, he chooses the price to offer.

Starting by examining the situation where in the first stage every generator submits all his available capacity. In the second stage of the game, each generator needs to decide if he wants to sell all his capacity, undercutting the prices offered by his competitors, or if he offers the highest possible price $\overline{P}$, exchanging quantity for price.

If a generator *i* decides to offer $\overline{P}$ his profit can be represented by equation (5.1).

$$\pi_i = \left(D - \sum_{j \neq i} K_j \right) . \left(\overline{P} - C_i \right) \tag{5.1}$$

However, if a generator i decides to undercut the price offered by the other generators in order to sell his full capacity, the profit function is represented by equation (5.2).

$$\pi_i = K_i . \left(P_i - C_i \right) \tag{5.2}$$

In equilibrium, generator i prefers to charge the maximum price if and only if $\left(D - \sum_{j \neq i} K_j \right) . \left(\overline{P} - C_i \right) \geq K_i . \left(P_i - C_i \right)$. Solving this equation for P_i, it is computed the upper bound $\left(\overline{P_i} \right)$ on P_i such that it is better for generator i not to undercut his competitors, equation (5.3).

$$P_i \leq \overline{P_i}$$
$$\overline{P_i} = \frac{\left(D - \sum_{j \neq i} K_j \right)}{K_i} . \left(\overline{P} - C_i \right) + C_i \tag{5.3}$$

This means that if, in order to sell his full capacity, a generator i has to undercut his rivals' offers by offering a price above $\left(\overline{P_i} \right)$, he should do so, otherwise, if the undercutting price is lower than $\left(\overline{P_i} \right)$ he should offer the maximum price, $\overline{P}$.

Assume that for all $i \neq M$, $\overline{P_i} < \overline{P_M}$ [21]. The properties of this Bertrand game with capacity constraints are now analysed. In this one-shot game, how would a generator

[21] Further, assume that when two or more generators offer the same price the algorithm applies a rationing rule. A commonly used rationing rule is the proportion of the available capacity, i.e., each generator offering the same price serves a proportion of the residual demand equal to the percentage he owns of the total capacity owned by the generators offering that price. Another efficient rationing criterion is the one in which the generators with the lowest marginal cost serve demand.

choose his offer price, given that the capacity is defined in the first stage? Next, Proposition 5.1 shows that every player can earn a risk-free profit in a possible outcome of the game. Further, Proposition 5.2 proves that this game has no Nash equilibrium. Then, Proposition 5.3 analyses the implications of the risk-free payoff on a generator's behaviour, by using the security-payoff criterion proposed by von Neumann and Morgenstern (1953)[22].

PROPOSITION 5.1: *In the one-shot pay-as-bid Bertrand game with capacity constraints, each generator* $i \neq M$ *earns the maximum security-payoff when generator M offers the maximum price and every i undercuts the upper bound of generator M; i.e.* $P_M = \overline{P}$ *and* $P_i = \overline{P_M} - \varepsilon$ *for every* $i \neq M$ *is the solution at which every generator earns his maximum security-payoff.*

PROOF: (a) Players $i \neq M$.

From equation (5.3), it follows that each generator may offer a price in the range $[\overline{P_i}, \overline{P}]$. Then, together with conditions $\sum_{i=a}^{N} K_i > D$ and $\sum_{i=a}^{N} K_i < D + K_s$ and the assumption that for all $i \neq M$ $\overline{P_i} < \overline{P_M}$, it follows that any generator $i \neq M$ can ensure that his full capacity is sold by offering in the range $\left[\overline{P_i}, \overline{P_M} - \varepsilon\right]$. Finally, by equation (5.2) it follows that by offering $\overline{P_M} - \varepsilon$ each generator *i* earns the maximum security-payoff.

(b) Player *M*.

22 Experimental evidence presented by van Huyck et al. (1990) and Cooper et al. (1990) supports this solution. They report experiments in which equilibrium converges to the security-payoff.

It follows from equation (5.3) that a generator M may offer a price in the range $\left[\overline{P_M}, \overline{P}\right]$. However, if as $\sum_{i=a}^{N} K_i > D$ and $\sum_{i=a}^{N} K_i < D + K_s$ and further assuming that for all $i \neq M$ $\overline{P_i} < \overline{P_M}$, then as any player i may offer a price in the range $\left[\overline{P_i}, \overline{P}\right]$, it follows that generator M earns the maximum security-payoff when he offers $\overline{P}$. Moreover, if any generator $i \neq M$ offers $\overline{P}$ as well, then M's payoff increases as he still sells at the price $\overline{P}$ and, depending on the rationing rules, he may increase, but not decrease, his output. Q.E.D.

PROPOSITION 5.2: *In the second-stage (after the capacity available is revealed) of the one-shot pay-as-bid Bertrand game with capacity constraints defined above there is no Nash equilibrium.*

PROOF: Proposition 5.1 shows that every generator $i \neq M$ offers in the range $\left[\overline{P_i}, \overline{P}\right]$, and that generator M offers in the range $\left[\overline{P_M}, \overline{P}\right]$, where $\overline{P_i} < \overline{P_M} < \overline{P}$, for all $i \neq M$. Then, since Proposition 5.1 shows that, for every generator $i \neq M$ the price $\overline{P_M} - \varepsilon$ gives these players the maximum security-payoff, it follows that they actually offer in the range $\left[\overline{P_M} - \varepsilon, \overline{P}\right]$ as the price $\overline{P_M} - \varepsilon$ dominates the offers in the range $\left[\overline{P_i}, \overline{P_M} - \varepsilon\right[$. Thus, the result is a matrix game played between each generator $i \neq M$ and generator M. Each player i may decide to offer the price that gives him the security payoff $\overline{P_M} - \varepsilon$ or to offer in the range $\left[\overline{P_M} + \varepsilon, \overline{P} - \varepsilon\right]$[23] that,

[23] The price $\overline{P_M}$ is dominated by the price $\overline{P_M} - \varepsilon$ as a generator i ensures to sell all his capacity without loosing any profit, since $\overline{P_M} \approx \overline{P_M} - \varepsilon$. Similarly, the price $\overline{P}$ is dominated by $\overline{P} - \varepsilon$, as

by equations (5.2) and Proposition 5.1, leads to a higher payoff than the security-payoff (but also to a higher risk). On the other hand, generator M may offer $\overline{P}$, which by Proposition 5.1 gives at least the security-payoff, or it may offer in the range $\left[\overline{P_M}, \overline{P}-\varepsilon\right]$ that, by equation (5.3), leads to increased payoffs if the player undercuts the price offered by at least one of his opponents (otherwise it leads to lower payoffs). Now, the best-response (BR) behaviour of the players is analysed:

a) If every generator $i \neq M$ offers $P_i = \overline{P_M} - \varepsilon$ then *M's* BR is $P_M = \overline{P}$, by Proposition 5.1.

b) If at least one generator $i \neq M$ offers $P_i \in \left[\overline{P_M} + \varepsilon, \overline{P} - \varepsilon\right]$ then *M's* BR is $P_M = P_i - \varepsilon$: by equation (5.3). The BR strategy of all other generators j such that $j \neq i \neq M$ is to offer $P_j = \overline{P_i} - \varepsilon$: by equation (5.2) and the possible offering range $\left[\overline{P_M} - \varepsilon, \overline{P}\right]$.

c) If generator M offers $P_M = \overline{P}$ then the BR of every player $i \neq M$ is to offer $P_i = \overline{P} - \varepsilon$: by equation (5.2) and the possible offering range $\left[\overline{P_M} - \varepsilon, \overline{P}\right]$.

d) If generator M offers $P_M \in \left[\overline{P_M}, \overline{P} - \varepsilon\right]$ then the BR of every player $i \neq M$ is to offer $P_i = P_M - \varepsilon$: by equation (5.2) and the possible offering range $\left[\overline{P_M} - \varepsilon, \overline{P}\right]$.

In this matrix game there is no Nash equilibrium as any of *M*'s BR to the strategic choices of every player i leads to at least one of these players to change his behaviour: by a) and c) or by b) and d). Similarly, a player i's BR to the strategic

shown in Proposition 5.1, since by the rationing rule player i sells less when generator M also offers $\overline{P}$, without compensating the quantity reduction by higher prices as $\overline{P} \approx \overline{P} - \varepsilon$.

behaviour of M and of all other players $j \neq i \neq M$ leads to changes in the behaviour of these players: by a) and c) or by b) and d). Q.E.D.

This simple game illustrates two very interesting characteristics of the security-payoff solution for the Bertrand competition with capacity constraints. a) When the market structure allows, generators tend to charge prices well above marginal cost. b) Capacity and marginal costs interact to define the players' pricing strategy. In the example above, generator M charges the highest price, but M is not necessarily the generator with the highest marginal cost.

Moreover, from equation (5.3) it is trivial to show that if any generator decides to reduce the available capacity in such a way that $\sum_i K_i = D$, then each generator, in equilibrium, can offer the maximum price, i.e., $P_a = \ldots = P_N = \overline{P}$. Thus, when will a generator withhold capacity? The answer is whenever this behaviour leads to higher profits. In this model, assuming the security-payoff solution, equation (5.4) describes the players' profits.

$$\begin{aligned} \pi_i &= K_i.\left(\overline{P_M} - C_i\right) \quad \text{for all } i \neq M \\ \pi_M &= \left(D - \sum_{i \neq M} K_i\right).\left(\overline{P} - C_i\right) \end{aligned} \tag{5.4}$$

Next, Proposition 5.3 shows that if a generator has the capability of withholding capacity it may be possible for this generator to increase the security-payoff that every generator in the industry can earn.

PROPOSITION 5.3: *Let K_i, $K_i^{'}$ stand for the full and reduced capacities of generator i. Further, let $\overline{P_M^{'}}$ represent the upper bound for generator M, when generator i decreases his capacity to $K_i^{'}$. Assuming the maximum security-payoff solution of the Bertrand game with capacity constraints, a player $i \neq M$ can increase his security-payoff by unilateral capacity withholding if and only if $\overline{P}$ is such that* $\left(D - \sum_{j \neq M} K_j\right) + \frac{C_M - C_i}{\overline{P} - C_M}.K_M < K_i^{'} < K_i$.

PROOF: By equation (5.2), a player $i \neq M$ can increase his security-payoff through capacity withholding if and only if $K_i^{'}.\left(\overline{P_M^{'}} - C_i\right) > K_i.\left(\overline{P_M} - C_i\right)$. Then by equation (5.3), $\overline{P_M^{'}}$ is a function of $\overline{P_M}$, and thus $\overline{P_M^{'}} = \overline{P_M} + \left(\overline{P} - C_M\right).\frac{Ki - K_i^{'}}{K_M}$. Replacing this expression in the profitability condition, one obtains a condition for profitability depending only on $\overline{P}$ and $\overline{P_M}$, i.e., $C_i - \overline{P_M} + \left(\overline{P} - C_M\right).\frac{K_i^{'}}{K_M} > 0$. Using equation (5.3), the profitability condition is derived as a function of $\overline{P}$ only, i.e., $K_i^{'} > \left(D - \sum_{j \neq M} K_j\right) + \frac{C_M - C_i}{\overline{P} - C_M}.K_M$. Finally, by definition of capacity withholding, it follows that $K_i^{'} < K_i$. Q.E.D.

Hence, Proposition 5.3 shows that, under the conditions defined, unilateral capacity withholding may increase the security-payoff earned by a player, leading to high prices (for the maximum security-payoff solution). These prices are higher than the ones predicted by the Bertrand model with full capacity available (as $\overline{P_M^{'}} > \overline{P_M}$).

Therefore, this analysis shows that profit margin is not a good indicator of abusive behaviour. Assume that a player i decides to withhold capacity, and that, as an outcome, a generator j increases his offer price such that he has the highest profit margin. In this case the market abuser is generator i but not generator j. Thus, the task of identifying individual market power is a challenging one, and *the regulator tends to look at capacity availability as the observable variable that indicates abusive behaviour by a certain generator* (as in bilateral contracts prices are private information).

Furthermore, this model also justifies why, in repeated games, it is important to consider learning behaviour when analysing the possible solutions of the game. In a repeated game, a player tries to learn the propensity of the generator M to undercut his prices. Thus, given the different risk propensities of the players different solutions can emerge. To which one of the possible solutions does the game converge? This is where learning theory enters by postulating that, in repeated games, a possible equilibrium solution may emerge from an iterative learning process, by trial and error, where a player tends to revise his pricing behaviour in order to maximise profit.

Finally, the aim of this simple Bertrand model is to motivate the analysis of the capacity withholding issue. Even in a situation where capacity is very tight and where each player owns only one type of technology, it may be possible for capacity withholding to be a profitable strategy. The model presented is clearly restrictive as it assumes that there is only one marginal generator, that a generator only owns one type of technology, and that there is a known maximum price. However, even this simple model does not have an equilibrium solution. The use of agent-based simulation, as will be presented in section 5.2, enables the researcher to develop models of the electricity industry that are more realistic than the ones designed by analytical game theory.

5.2 The AES-British Energy Case

This section presents the experiments used to analyse the capability of AES and (or) BE to influence the wholesale electricity prices.

Table 4.9 presents the market-shares of the eight largest generators in the E&W electricity market (in 2000). As shown with the simple example in the previous section, a generator's pricing behaviour is a function not only of the technologies owned (the marginal costs), but also of his market share. Hence, it is important to look at the relative market share of each generator, in each type of technology, in order to understand its *strategic position.* Thus BE had 54% of the Nuclear generation capacity installed in E&W (and 4.9% of the shoulder capacity), while AES owned both shoulder (10.1% of shoulder capacity) and peak plant (6.8% of peak capacity). Table 4.10 presents the market shares for the retail supply business (in 2000). By only looking at market shares, it is noticeable that the supply business is less concentrated than the generation business. The biggest player is London Electricity with 12% of market share.

A detailed agent-based model was applied to the full system of E&W as it existed in 2000. Some small gensets were aggregated and overall the model was specified with 80 gensets, owned by 24 generators who sold power to 13 suppliers.

In the simulations presented the total capacity in the system was 59 GW. Further, these simulations use three different demand profiles for winter (with peaks 49 GW, 48 GW and 45 GW) and one for summer (with three maintenance scenarios 14 GW, 12 GW and 10 GW). Figure 5.1 presents the demand profiles used in these simulations. The experiments in this chapter assume that each supplier exhibits a demand forecasting error of 10%.

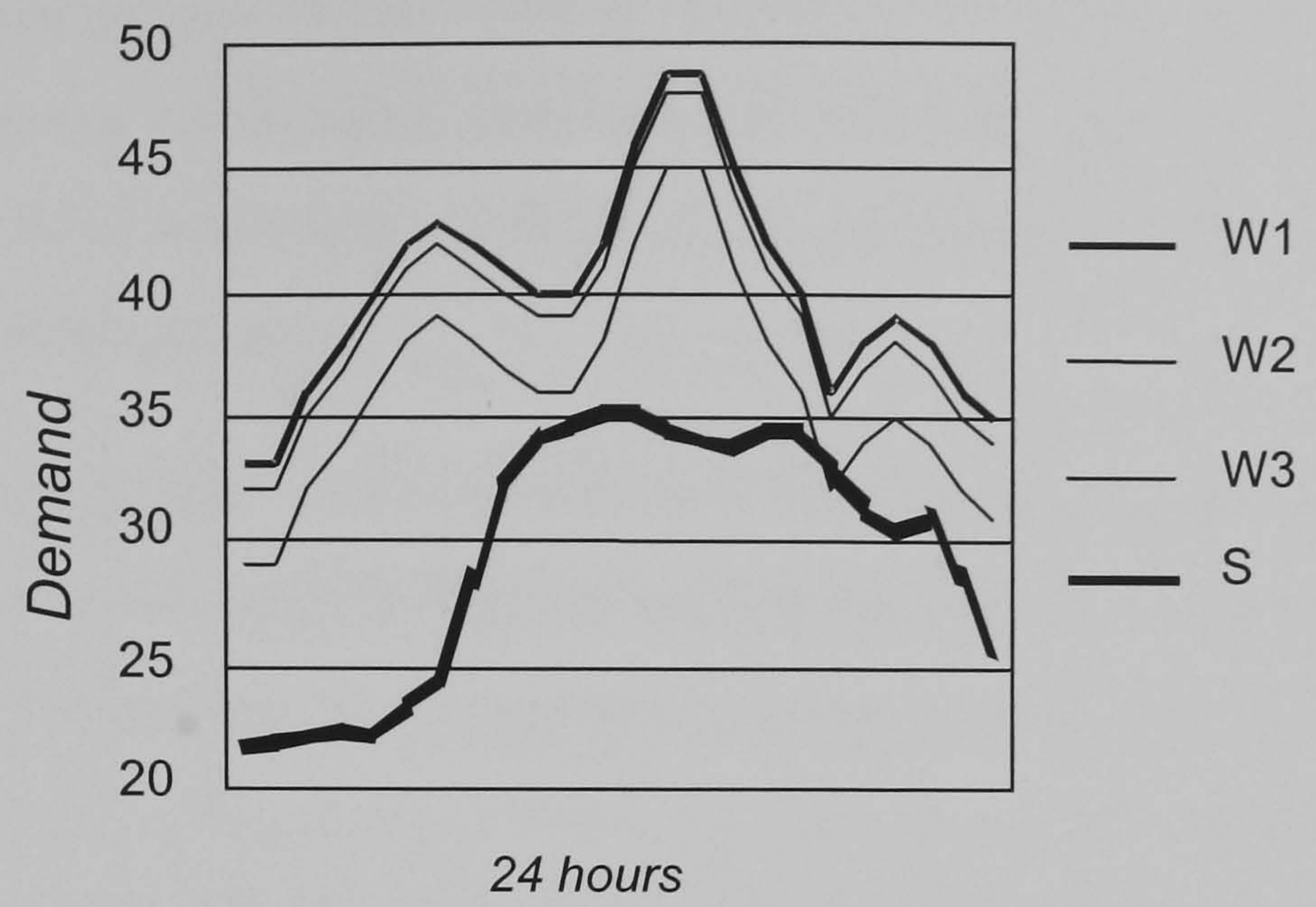


FIGURE 5.1 Demand profiles. Lines W1, W2 and W3 stand for winter demand with a peak of 49, 48 and 45 GW respectively. Line S stands for summer demand with a peak of 35 GW

Active trading between generators and suppliers was simulated as a multistage process through a PX (that works as a proxy for all the forward bilateral markets) and a BM. All the comparative experiments have simulated 50 iterations (trading days) of the algorithm and the results were averaged over the final 10 days. In most cases, the evolution of prices seemed to settle around a stable position after the initial learning, by about 50 iterations.

Six different *Strategies* were simulated for AES and (or) BE in relation to making capacity available under six different *Scenarios*. In strategy ST1, AES and BE offer their full capacity. In strategy ST2, AES saves capacity for the BM (480 MW). In strategy ST3, BE saves capacity for the BM (490 MW). In strategy ST4, AES saves capacity for the BM (480 MW) and closes a Drax genset (645 MW). In strategy ST5 BE saves capacity for the BM (490 MW) and closes an Eggborough genset (480 MW). Finally, in strategy ST6, AES saves capacity for the BM and closes a genset, and at the same time BE saves capacity for the BM and closes a genset, i.e., strategies ST4 and ST5 combined.

Further, these simulations tested several different scenarios. Specifically, three winter day scenarios with peak demands of 49 GW, 48 GW and 45 GW, respectively; and three summer demand scenarios, maintenance outage of 14 GW, 12 GW and 10 GW, respectively.

Each one of these scenarios was repeated twice with small differences (that is, with different random impacts or with different maintenance levels for some plants). This analysis starts by comparing the average prices of the six strategies (under the twelve simulations). As expected by the theoretical Bertrand model with constraints, in these simulations the average PX price and *SBP* are higher than the marginal costs (even with full capacity available).

A first conclusion of this analysis was that E&W wholesale electricity market had a structure that would give rise to prices above marginal costs, and therefore non-competitive. This implies the existence of market power in the *traditional sense* (capability to rise prices above marginal costs) that results from the overall structure of the industry. In this sense, it is the joint behaviour of all players (profit maximising without explicit collusion), in an imperfect market, that leads to higher prices. Therefore, it is not possible to say that any individual player is abusing his market power. If, however, a player seeks to alter the market structure by deliberately withholding capacity, then the regulatory perspective has clearer grounds for identifying individual market power abuse.

The capability of AES and BE to manipulate market prices through capacity withholding can be evaluated by analysing the market's performance when the players withhold capacity with its performance when full capacity is available. As shown in the previous section, this implies comparing the prices under strategies ST2, ST3, ST4, ST5 and ST6 with the prices under strategy ST1.

Unfortunately, this task is not as straightforward as in the Bertrand stylised analysis. The analysis of the NETA computational model needs to consider the possible presence of multiple equilibria and the learning effects. The analysis of the results in these experiments uses t-statistics (for pooled samples) to compare the different simulations. If the t-statistic associated with a certain strategy (ST2, ST3, ST4, ST5 or ST6) is higher than 2.2 (the critical value for 5% significance level with 11 degrees of freedom) then that strategy has a significant impact on prices or profits[24], otherwise there is no statistical evidence that the particular strategy has an impact on prices and (or) profits. Table 5.1 presents the t-statistics for the PX price (*PXP*), *SSP* and *SBP*.

TABLE 5.1: t-statistics for market prices

	Strategies				
	ST2	ST3	ST4	ST5	ST6
PXP	1.12	2.21	2.61	2.37	3.55
SSP	0.23	0.83	1.92	0.26	1.7
SBP	-0.15	0.95	0.75	1.98	1.5

The experiments show that AES and BE have enough market power to increase the *PXP* to levels above the normal prices resulting from the full capacity strategy ST1. Nevertheless, these players cannot influence the *SSP* or the *SBP*. The BM prices tend to be more dependent on random impacts and path dependencies than on the manipulation attempts by AES and (or) BE. Therefore, there is statistical evidence, through these simulations, that AES and BE can behave as price makers. However, is this price making ability synonymous of higher profits? Can they profit from capacity withholding?

In order to analyse the profitability of each strategy (ST2, ST3, ST4, ST5 or ST6), AES and BE's profit contributions under each one of these strategies are compared with their profit contributions under the full capacity strategy ST1. Again, the

[24] The Kolmogorov-Smirnov test for normality cannot reject the hypothesis of normality for the sample distribution of the differences, for both prices and profits. The Wilcoxon nonparametric test

profitability issue is analysed using t-statistics for the pooled samples. This analysis compares the changes in profit for each strategy (when compared with strategy ST1) under each scenario.

The t-statistics in Table 5.2 show that only when both players simultaneously withdraw capacity from the market (and save capacity for the BM) can they profit from price manipulation. These results seem to suggest that profitable price manipulation appears to be harder than expected by the standard market power literature in electricity markets.

TABLE 5.2: t-statistics for AES' s and BE' s profits

	Strategies				
	ST2	ST3	ST4	ST5	ST6
AES	-0.09	1.77	1.31	1.61	2.29
BE	-1.03	0.32	0.58	1.57	3.04

5.3 Conclusions

This chapter presents a stylised model of a market for electricity (modelled as a Bertrand game with capacity constraints), which is used to show that even when all generators offer their full capacity, and even if the total capacity available is higher than the total demand, prices can be much higher than the marginal costs (Proposition 5.1). Further Proposition 5.2 shows that unilateral capacity withholding may lead to increased profits and misevaluation of market power abuse in certain circumstances.

---

also did also corroborate the overall conclusions for both prices and profits.

The research question addressed in this chapter is the evaluation of the conduct of an individual generator. In the case under study, Ofgem (the E&W regulator) were worried about the possibility of AES and BE having the capability to profitably manipulate market prices.

In order to test if AES and BE could manipulate the bilateral markets or the balancing mechanism, and (or) the interactions between them, several scenarios were simulated aiming to study these generators' capability to increase prices.

The literature on market power generally evaluates the competitiveness of a market by comparing the prevailing prices with the expected marginal costs. One generalisable contribution of this work is to reinforce the perspective that, in an imperfect market, even when all the generators offer all their capacity, individual profit maximisation without overt collusion will result in average prices substantially higher than the marginal costs. This shows that, even if all players behave competitively, the structure of the England and Wales electricity market will be such that prices will significantly tend to rise above marginal costs, even without considering other technical constraints such as ramping rates and transmission. Given the market structure, it is the competitive outcome where all players behave as profit maximisers but non-strategically in the market, which becomes the crucial point of reference from which to evaluate conduct. This is where a model is required and the agent-based simulations undertaken here did provide a rich basis for such analysis, albeit statistically.

Regarding the specific questions raised by the Competition Commission, the computational experiments show that:

1. BE alone, and BE simultaneously with AES, may significantly increase the Power Exchange prices.
2. AES alone cannot significantly influence the wholesale Power Exchange prices.

3. The System Sell Price and the System Buy Price cannot be significantly influenced by the strategies implemented by AES and (or) BE.

4. Only when acting simultaneously may the firms profit significantly from capacity and price manipulation.

The Competition Commission concluded that it would *not* be against the public interest for the market to operate without modifications of the licence provisions of AES and BE to include a clause seeking to preclude Market Abuse, mainly on the basis of the awkward principle of this as a regulatory instrument. (Competition law was considered sufficient, by itself, to deal with market abuse, if it did become apparent.)

Overall, the agent-based simulation technique enabled substantial insight into the behaviour of the new wholesale electricity market in 2000, before its introduction in 2001. This technique enabled the modelling of complex adaptive behaviour in an environment with possible multiple equilibria, with heterogeneous agents and price uncertainty. This shows that models having the functional capability to represent the type of learning occurring in real electricity markets can capture the processes underlying the equilibrium selection and behaviour coordination in complex electricity markets.

Furthermore, as the experiments in this chapter assume a constant demand profile for each scenario of the game, the simulation may be perceived as an iterative approach to compute the solution of the one-shot game. However, since a player learns the rewards associated with his policy by interacting with other players (and discounted from past experience), these repetitions of the game may change the perceptions an agent has of the others' behaviour. Thus, the NETA learning algorithm captures this evolutionary process opening the door to implicit collusion.

However, it is plausible that, since this learning process occurs in a stable environment where agents iteratively adapt their behaviour to specific conditions, this type of modelling over-estimates the potential for strategic coordination. In real markets the players have to learn in an environment where demand, fuel costs, and transmission constraints reshape the system continuously, which makes behaviour coordination harder to achieve.

As observed in Chapter 4, the value of a plant is different under different types of portfolio. Further, there is evidence that in the UK, in California or in the Iberian market, companies are trading generation plants. Would it be possible to develop a theoretical analysis and a computational model to explain and capture this behaviour?

The trading model developed in Chapter 4 cannot handle a setting where players are able to trade plants, as the learning policies are portfolio dependent. Thus, in order for a player to learn the value of a given plant he currently owns he would have to sell it, in order to evaluate the portfolio without it. Likewise, in order to evaluate its contribution under his existing portfolio he would have to buy a plant that is currently not working.

Another reason not to use the simulation approach presented in Chapter 4 to study the strategic management of portfolio of plants is that an analytical approach is necessary in order to assure that generalisation is possible. In addition, another problem with the NETA model to study evolutionary problems is that, although attractive from a theoretical point of view, models of learning have the shortcoming of slow speed of convergence that turns them difficult to use within an evolutionary setting.

Moreover, as shown in Chapter 7, a model of learning and adaptation that captures and shapes the trajectory toward equilibrium is required to simulate plant trading in electricity markets, as plant trading occurs in off-equilibrium states. However, as presented in Chapter 3, models of reinforcement learning cannot simulate this type of game, as one needs to capture the behaviour of a player that is able to infer a model of the environment, and to evaluate a given action, without playing it. A model that embodies these characteristics is the Finite Automata Dynamic Game (FADG) presented in Chapter 6, which Chapter 7 applies to the plant trading issue.

CHAPTER 6

LEARNING AND ADAPTATION IN FINITE AUTOMATA GAMES

The deductive equilibrium methods such as the Nash equilibrium and the rational expectations paradigm are powerful tools for analysing industries where there are strategic interdependences between players. However, these methods do not explain the process by which decision makers acquire equilibrium beliefs, failing to determine a unique equilibrium solution in many games, and hence failing to predict or prescribe *rational behaviour* (see van Huyck et al., 1990; Weibull, 1995; Samuelson, 1997; Fudenberg and Levine, 1998). Further, these methods also fail to analyse the uncertainty introduced in a game by boundedly rational behaviour.

This argument emphasises the need to model opponent's behaviour, which is formalised by Rubinstein (1986, 1998) using deterministic finite automata (see Chapter 3). Further, as justified in Chapter 5, a reinforcement learning model cannot be used in a environment where players are able to trade assets, as learning policies are portfolio dependent and the structure of the game evolves over time.

This chapter uses learning and adaptation to capture the dynamics issue, showing that if players use rules of behaviour learned from experience, even if they may change these rules, automata theory can capture the dynamics of this process.

The FADG challenges a previous claim by Rubinstein (1998) that automata can only be used to model one-shot games, and that a player cannot be allowed to change his automaton[25]. This claim is certainly true in a world where a player can change his automaton at every stage and has perfect information regarding the automata set used by his opponents. However, in a world where people and companies do use rules of behaviour, and where the information collection and the adaptation processes are costly, it may be possible and useful to model an automata game where the players may be able to learn and adapt. This is one of the contributions of the research here developed.

Thus, this chapter develops a new framework to model learning and adaptation in N-player extensive games of incomplete information, the Finite Automata Dynamic Game (FADG). The FADG models the set of possible automata as endogenous variables, capturing the process through which a model of other players' behaviour is learned, and enabling the analysis of the process from which behaviour emerges. The FADG is a game with incomplete information, played by automata, where each player infers a model of the environment using an identification algorithm, Quasi-Perfect-rationality (QPR), and may change his automaton during a game by using an adaptation algorithm, Adaptive Best-Response (ABR).

Another contribution comes from the QPR algorithm that enables a player to forecast the outcome of any action off-equilibrium path without experimenting it, i.e., the FADG player implicitly explores the environment without having to learn only by trial and error. Please note that the traditional reinforcement learning models (e.g. Sutton, 1988; and Bertsekas and Tsitsiklis, 1996), assume that a player can learn by trial and error, even by choosing actions off-equilibrium path (exploring the environment). However, in certain games, the consequences of these actions may be very important, even irreversible, and therefore a player cannot use reinforcement learning. Hence, the QPR algorithm allows a player to learn without experiment all his actions, avoiding potentially costly choices, and learning in a more plausible fashion.

Before proceeding, this paragraph presents a brief sketch of the FADG. At each stage of the automata game, a player holds a certain model of the environment (the perceived residual product automaton). This perceived residual product automaton represents a model of the environment's aggregated behaviour at a certain stage (note that this behaviour is a function of the automata used by other players). Given this model and the information collected by interacting with the environment, i.e., the payoffs received, a player computes the plausibility of the inferred model. If this plausibility is high enough for the player to believe that the model represents accurately how the environment will behave in the future, the agent may decide to change his behaviour in order to increase his expected payoff (profits or utilities, depending on the specific game modelled). Then, a player computes the best response strategy to play against the inferred model. At this point, a player has to decide between keeping the same rule of behaviour (the same automaton) or changing it to a new one, given the equilibrium computed by using the best response algorithm. The literature in automata theory (e.g., Rubinstein, 1986; Gilboa, 1988) seems to recommend that the player should adopt the best-response automaton. However, in a world with multi-equilibria, and possibly conflicting ones, the best response behaviour may lead the player to a second best (even if the model of the environment is correct). So instead a player may decide to keep the same automaton,

[25] Previously Carmel and Markovitch's (1996, 1999) work also partially challenges this claim by allowing a player to change his automaton as he interacts with his opponent (which is assumed static).

imposing his behaviour to his opponents in an attempt to gain credibility and to transfer the adaptation cost to the other players, i.e., the player behaves with active inertia.

Hence, in a world where players exhibit behavioural inertia, due to imperfect information or strategic "stubbornness" (in this game, the players do not change their behaviour in each stage of the game), it makes sense to model the dynamics behind the emergence of automata behaviour. This model may be able to explain the emergence of certain types of behaviour as social norms without imposing the possible set of automata exogenously.

This last point is noteworthy. In an FADG, each automaton is endogenous, as it is the product automaton of the automata playing at a certain time. This means that the model does not constrain the behaviour of the players to certain patterns, and that this is an emergent behaviour of the game.

At an analytical level, this chapter shows that the best-response behaviour is not a sufficient condition for rational behaviour, and that behavioural inertia is a necessary condition for rational behaviour in a world where the presence of boundedly rational players generates uncertainty.

The chapter proceeds by introducing the concept of FADG, by presenting the QPR and ABR algorithms, and by discussing the equilibrium properties of the FADG. It is important to note that, although presented in two distinct sections, there is a strong link between the QPR and ABR algorithms, as each player adapts to his perceptions, and the rules of behaviour actually used by a player influence the outcomes received and the perception a player holds of the environment.

Finally, the chapter ends with two sets of simulations that use the FADG and with a summary of the findings. The first simulation uses the Keynesian coordination game and shows that the FADG framework can replicate the way people play these games. The interesting result of these experiments is to show that when players have high forecasting error the results tend to approximate better the behaviour of the games played by people. The second set of experiments presents the pie-game. These simulations show that if all the players are able to learn and adapt, the game may evolve to the perfect-equilibrium solution. However, the capability of the players to infer correct models of the environment is crucial for the stability of the equilibrium to which the game converged.

6.1. Finite Automata Dynamic Game

The literature defines the automata game in the strategic form (Rubinstein, 1988). However, this assumes a game of complete information where the player knows the automaton used by his opponent and tries to compute the best-response; or it assumes that both players know the set of automata available, and each one chooses the automaton that maximises his payoff in a repeated game, knowing that his opponent also knows, and so on ...; or finally, it may assume that a player always holds the same automaton and the other attempts to infer this automaton by playing repeatedly against him.

However, in games that define the automata as co-evolutionary rules of behaviour, the players do not know the payoff functions, they have to learn them by playing the game. More, these players play the same game repeatedly and, since by definition of bounded rationality there is no reason to assume common knowledge of rationality, each one of them is required to infer a model of his opponents. Therefore, it seems that, in this setting, the extensive form is a better formalisation of a co-evolutionary automata game.

An automata game in the extensive form is a 5-tuple $G = \left(\mathrm{N}, \left\{ \mathrm{Z}^i \right\}_{i=1}^{N}, \left\{ u^i \right\}_{i=1}^{N}, \left\{ Q^i \right\}_{i=1}^{N}, \left\{ \Sigma^i \right\}_{i=1}^{N} \right)$. N denotes the number of players. Z^i represents a finite non-empty set of possible outcomes of the game, and each $z^i \in Z^i$ is a function of the actions of each player, $z^i = z\left(a^i, a^{-i}\right)$, where $a^i \in \Sigma^i$ represents an action of player i and $a^{-i} \in \Sigma^{-i}$ represents his opponents' actions. The outcomes of the game represent the information received by each player at the end of every stage. This information, or outcome, is a function of the actions of each player in the stage game, and it is different for each one of the players, as each one only knows the outcome of his own actions. $u^i = u\left(z^i\right)$ represents the utility function of player i, i.e., it is the payoff a player i perceives to have received from his action, given the perceived outcome. Q^i stands for a finite non-empty set of internal states of player i. Σ^i is a non-empty set of all possible actions of player i. The automata game G is an extensive form game where each player evolves a certain decision rule that may change at a certain iteration of the game. This decision rule, his automaton A^i, defines how a player reacts to the outcomes received from the environment. Given that information is incomplete and costly, and given the behaviour inertia later defined, a player may use the same automaton repeatedly, even if his behaviour is not optimal. This is one of the major characteristics of evolutionary games where a best-response dynamics with inertia defines the adaptation or "selection" mechanism (Samuelson, 1997: 242). Finally, for reasons of notational simplicity this chapter drops the time indices (however, these are included when needed for clarity purposes).

In the extensive-form automata game G, each player i can be described by a *characteristic*[26] $\mathrm{K}\left(A^i, P^i\right)$, consisting of a finite adaptation automaton A^i that specifies a players' behaviour at a certain time, and a perceived residual product

[26] Following Samuelson (1997: 241).

automaton, P^i, that specifies how a player models the environment, i.e., his opponents. A finite automaton $A^i = \left(Q^i, q_0^i, \Sigma^i, \delta^i, \lambda^i\right)$ is a 5-tuple that represents the automaton used by the player i, and where a player i aims to maximise his payoff when playing against $\{A^j\}_{j \neq i}$. Q^i is a finite non-empty set of internal states; q_0^i is the initial internal state; Σ^i is the set of all the possible actions; δ^i is a transition function $\left(\delta^i : Q^i \times Z^i \rightarrow Q^i\right)$ and λ^i is a behavioural function $\left(\lambda^i : Q^i \rightarrow \Sigma^i\right)$ associating an action to each possible internal state. The perceived residual product automaton $P^i = \left(Q^{pi}, q_0^{pi}, \Sigma^{pi}, \delta^{pi}, \lambda^{pi}\right)$ represents the model the player holds of the environment at a certain stage.

A player, at a certain stage of the game, may decide to keep or change his automaton (although due to the inertia factor this change seldom happens). At stage one after the automaton was changed, each player i plays $\lambda^i\left(q_0^i\right)$. At a stage $t \geq 1$, after each player executing his actions with an outcome $z_t^i = \lambda^i\left(a^i, a^{-i}\right)$, each automaton A^i moves from the state q_t^i to the state $\delta^i\left(q_t^i, z_t^i\right)$. Then each player i chooses a new move $\lambda^i\left(q_{t+1}^i\right)$.

DEFINITION 6.1 (FADG)[27]: *A Finite Automata Dynamic Game is an N-player discrete-time game, in the extensive form, with incomplete information, where each player i faces a sequential decision process* $\Pi^i = \left(A^i, P^i, u^i, V^i\right)$ *such that:*

(i) $A^i = \left(Q^i, q_0^i, \Sigma^i, \delta^i, \lambda^i\right)$ *represents player i's automaton.*

(ii) $P^i = \left(Q^{pi}, q_0^{pi}, \Sigma^{pi}, \delta^{pi}, \lambda^{pi}\right)$ *represents player i's perception of the residual product automaton of all N-i players, while* $M^i = \left(Q^{mi}, q_0^{mi}, \Sigma^{mi}, \delta^{mi}, \lambda^{mi}\right)$ *is the true*

[27]The concept of sequential decision process follows Karp and Held (1967).

residual product automaton of the game, representing the residual product automaton of all N-i players.

(iii) $u^i = u(z^i)$ *represents the utility function of player i, i.e., it is the payoff a player i perceives to have received from his action, given the perceived outcome* z^i. V^i *represents the value at stage 1 of the automaton used by i.*

(iv) W is the true product automaton of the game. At a certain stage, W is a 5-tuple $W=(Q,q_0,\Sigma,\delta,\lambda)$ *that defines how the environment behaves, where* $Q=Q^1\times Q^2\times\ldots\times Q^N$, $q_0=q_0^1\times q_0^2\times\ldots\times q_0^N$, $\Sigma=\Sigma^1\times\Sigma^2\times\ldots\times\Sigma^N$, $\delta:Q\times\Sigma\rightarrow Q'$, *and* $\lambda:Q\rightarrow\Sigma$.

(v) The objective of each player is to maximise V^i *by choosing a behavioural function* λ^i *and a transition* δ^i.

(vi) The information set of each player, at each stage, contains the current perceived residual product automaton, P^i, *and the new data arriving from the interactions with the environment,* D^i, *(which represents the perceived outcomes of the player's actions in the last path of his automaton,* A^i*).*

*(vii) An identification algorithm (Quasi-*Perfect-rationality*) updates the player's perceived residual product automaton,* P^i, *transforming the stream of data (actions and perceived outcomes) into a new* P^i.

(viii) The Adaptation algorithm (Adaptive Best-Response) transforms the stream of data (actions and perceived outcomes) into a new automaton, A^i.

Table 6.1 represents the FADG algorithm.

TABLE 6.1: Finite Automata Dynamic Game algorithm

Algorithm FADG (G)

G: Definition of the finite automata game in the extensive-form

N: Number of Players

$\Sigma = \{\Sigma^1, \ldots, \Sigma^N\}$: Set of possible actions of each player $i = 1, \ldots, N$

$A_0 = \{A_0^1, \ldots, A_0^N\}$: Initial set of automata

ρ_i : Discount factor for player i

w^i: Forecasting error: probability of a player to estimate the wrong outcome for the actions off-equilibrium path

D^i: Data collected by player i during a cycle of interactions

V^i: Value of the automaton used by a player i

Furthermore, consider the following notation:

a_t^i : An action of player i at stage t

u_t^i : The reward of player i at stage t

z_t^i : The outcome of the game at stage t

While last iteration is not reached

Simulate gaming in a *cycle* of interactions

Each agent makes a move given his current automaton

$$a_t^i = \lambda^i\left(q_t^i\right)$$

Compute the new state and the reward of each agent

$$q_{t+1}^i = \delta^i\left(q_t^i, z_t^i\right)$$

$$u_t^i = u^i\left(z_t^i\right)$$

For each agent i in the game approximate the value V^* of A^i:

Approximate the value of every state q_t^i in A^i

$$V^*\left(q_t^i\right) = u_t^i\left(z_t^i\right) + \rho_i V_{t+1}^*\left(\delta^i\left(q_t^i, a_t^i\right)\right)$$

For each agent i in the game call the identification algorithm:

$$P_{t+1}^i = QPR\left(D_t^i, w_t^i\right)$$

For each player i compute the new adaptation algorithm

$$A_{t+1}^i = ABR\left(A_t^i, P_t^i, V_t^i, \rho_i\right)$$

First, note that in this chapter only the discounted game is analysed. However, it would also be possible to use other measures of utility. Second, the data, D^i, used to update the perceived residual product automaton, P^i, refers to the last path in the automaton A^i, as it represents the last configuration to which the product automaton W converged. Third, the forecasting error, w^i, represents a way of incorporating Selten's (1975) "trembles" into a player's perceptions of the environment. Moreover, an algorithm of reinforcement learning (see Chapter 3 for an analysis of the learning process) approximates the value function V^*.

Finally, Abreu and Rubinstein (1988) and Gilboa (1988) show that the equilibrium of the normal form automata games with an infinite number of stages represents a cycle in a player's automaton. However, this analysis assumes that a perfectly rational player chooses the best response automaton against the opponent's choice (in a game where there is a set of possible automata that a player can chose). In a game in the extensive-form of incomplete information, where a player does not have a set of possible automata to choose from (he has to build his automaton using a best-response algorithm), a player may be able to evaluate his current automaton and decide to re-build it, i.e., the set of automata is endogenous. Thus, when the game reaches a cycle each player is able to evaluate his own automaton, deciding how to adapt to his new perceptions.

6.2. The Identification Algorithm

In order to model rational behaviour from limited experience in environments where information is incomplete, the Quasi-perfect-rationality (QPR) algorithm was developed, which builds upon Gold's (1978) and Angluin's (1987) algorithms for automata inference.

A player *i* infers a residual product automaton M^i. This learner has a set of possible actions Σ^i that are the input for the residual product automaton. In addition, the residual product automaton M^i has a set of outcomes Σ^{mi}. The learner *i* maintains an *observation table* (*S, E, T*) which he uses to infer M^i. This table represents the interactions with the residual product automaton M^i, where $S \subseteq \Sigma^{i*}$ is a non-empty finite pre-fixed closed set of strings, $E \subseteq \Sigma^{i*}$ is a nonempty suffix-closed set of strings called tests[28], and *T* is a finite two-dimensional table with one row for each element of the set $S \cup S\Sigma^i$, where $S\Sigma^i = \left\{ s\sigma / s \in S, \sigma \in \Sigma^i \right\}$, and one column for each element of *E*.

A player *i* learns an observation table *T* that is complete and consistent.

An observation table is said to be complete if, for each element $\phi^i \in S\Sigma^i$, there is an observation $s \in S$ such that $row(\phi^i) = row(s)$[29]. Thus, the completeness requirement implies that the player has a forecast for every action in every state of the automaton model.

An observation table is said to be consistent if for any two elements $s_1, s_2 \in S$, such that, $row(s_1) = row(s_2)$, and for all actions $a \in \Sigma^i$, $row(s_1 a) = row(s_2 a)$. Therefore, the consistency requirement implies that an agent has a correct automaton model, i.e., a model that does not forecast different transitions for the same action at a certain state.

---

[28] A set is prefix-closed if and only if every prefix, of every member of the set, belongs to the set. A set is suffix-closed if and only if every suffix, of every member of the set, belongs to the set. Note that the empty string belongs to both the prefix-closed and the suffix-closed sets.

[29] The row operator transforms the possible sequence of outcomes of the game into expected new outcomes, i.e., $row: (S \cup S\Sigma) \rightarrow (S \cup S\Sigma)$. Please note that the row operator allows a player to forecast the possible outcomes of his actions in information sets off the equilibrium path without having to try them. This represents the main improvement of this model when compared to reinforcement learning.

Next, Proposition 6.1 shows that the ability of a player to infer a complete and consistent model of residual product automaton is a necessary condition for rational behaviour. However, the proof of this proposition requires a definition of rational behaviour.

A *definitive* definition of rational behaviour is not available still. Von Neumann and Morgenstern (1953: 9) argue that a rational agent *attempts* to maximise his utility; van Damme (1991: 1) also defends this "normative theory" and maintains that the goal of game theory is to provide a solution for every game, given *a* definition of "rational behaviour". However, as pointed out by Samuelson (1997: 3), the problem with this approach to game theory is the "self-reference" created by the definition of "rational behaviour". Hence, having acknowledged this limitation, this chapter adopts the meaning of rational behaviour presented in Definition 6.2.

DEFINITION 6.2: *Rational behaviour is the capability of a player to adapt his behaviour in order to achieve the highest payoff in a multi-stage game. More formally, at each stage t, by changing* A_t^i*, a rational player aims to maximise*

$$V_t^i = \sum_{j=t}^{+\infty} \rho_i^j E_t\left(u_j^i\left(z_j^i\right) / A_t^i\right).$$

Next, this chapter shows that this rational behaviour is not easy to achieve, given the uncertainty regarding other players' behaviour. In a world of boundedly rational players, the behaviour of other players is uncertain as it depends on their learning and adaptation processes.

A rational player tries to maximise the present value of all the payoffs received during his lifetime. However, as can be understood through Definition 6.2, his ability to do so depends on his expectations. The residual product automaton P^i contains a model of these expectations. This means that rational behaviour in a multi-stage game of incomplete information is dependent not only on the ability of the player to compute his best-response to a given environment, which is approximated by the reinforcement learning equation $V^*\left(q_t^i\right)=u_t^i\left(z_t^i\right)+\rho_i V_{t+1}^*\left(\delta^i\left(q_t^i,a_t^i\right)\right)$, but also on his ability to infer a plausible P^i. In this case, as shown in Proposition 6.1, the capability of a player to infer a complete and consistent P^i is a necessary condition for rational behaviour. Further, as shown in the next section, and as suggested by Samuelson (1997: 242), a player that does not find his model (P^i) a plausible representation of the environment may decide to keep the same behaviour A^i.

PROPOSITION 6.1: *A necessary condition for rational behaviour is the capability of a player to infer a complete and consistent model (P^i) of the environment.*

PROOF: Let equation $V^*\left(q_t^i\right)=u_t^i\left(z_t^i\right)+\rho_i V_{t+1}^*\left(\delta^i\left(q_t^i,a_t^i\right)\right)$ represent the best-response function. This equation approximates the best response behaviour, given P^i.

Completeness: Assume that the model inferred by a player i is not complete, i.e., for some states $q_o^{pi}\in Q^{pi}$ this player has no forecast for the output of some of his actions. Assume also that a_t^i is the action to which player i holds no forecast. Then it is impossible for a player i to compute the transition function $q_{t+1}^{pi}=\delta^{pi}\left(q_{o,t}^{pi},a_t^i\right)$, and therefore it is not possible to compute the best-response action. It follows that this player cannot exhibit rational behaviour.

Consistency: Assume that the model inferred by a player i is not consistent i.e., for some states $q_o^{pi}\in Q^{pi}$ the same action a_t^i has different transitions, e.g.,

$q^{pi}_{1,t+1} = \delta^{pi}\left(q^{pi}_{o,t}, a^{i}_{t}\right)$ and $q^{pi}_{2,t+1} = \delta^{pi}\left(q^{pi}_{o,t}, a^{i}_{t}\right)$. Again, the best-response function is ill defined and this player cannot exhibit rational behaviour[30]. *Q.E.D.*

Thus, a QPR player has the capability of learning the residual product automaton consistent with the data received. Furthermore, the player uses counter-examples resulting from interacting with the opponents to revise the model.

Hence, the QPR player implicitly explores the environment, forecasting the future outcomes of the residual product automaton without having to try every action. Table T (presented in Table 6.2) is a forecasting table. Given a succession of actions $a^{i}_{t}, a^{i}_{t+1}, \ldots, a^{i}_{t+k}$, the player builds a complete and consistent model by forecasting $z^{i}_{t+k+1}\left(a^{i}_{t+k+1}, a^{-i}_{t+k+1}\right), z^{i}_{t+k+2}\left(a^{i}_{t+k+2}, a^{-i}_{t+k+2}\right), \ldots$ for every possible $a^{i}_{t+k+1}, a^{i}_{t+k+2}, \ldots$ and so on, until he builds a complete and consistent model.

Table 6.2 presents the QPR identification algorithm. This algorithm differs from Angluin's algorithm as it considers a player that faces an environment with incomplete information, where new experiences are hard to get, i.e., where exploration of off-equilibrium actions may be extremely expensive.

The QPR algorithm differs from Carmel and Markovitch's (1996) algorithm since it assumes that a player has learning capabilities that enable him to forecast the possible behaviour of the environment with a certain prediction error, enabling him to implicitly explore the possible behaviour of the residual product automaton without experimenting it. This is the main contribution of this algorithm, which avoids the trial and error learning with actions off-equilibrium path that a minimum rational player would not play.

[30] Note that the value of a state in the automaton A^{i} depends on the transitions in the automaton P^{i}. There is a one-to-one correspondence between these automata, which section 6.3 explains by analysing the adaptation algorithm.

TABLE 6.2: QPR identification algorithm

Algorithm $QPR(D^i, w^i)$

D^i: Data collected by player i during a cycle of interactions

P^i: The current residual product automaton model consistent with D^i

Σ^i: Set of possible actions of player i

$\phi^i = (s, \sigma)$: New example of the product automaton's behaviour

ε: Empty string

w^i: Forecasting error: probability that a player commits a mistake in estimating the outcome for the actions off-equilibrium path

S: Non-empty finite pre-fixed closed set of strings

E: Non-empty finite suffix closed set of strings, called tests

T: Finite two dimensional table

Initialise (S, E, T)

$S \equiv$ All prefixes of D^i

$\forall s \in S$ and $\sigma \in \Sigma^i$ if $s\sigma \notin S$ then $S\Sigma^i_{k+1} = S\Sigma^i_k \cup \{s\sigma\}$

$E = \{\varepsilon\}$

$\forall s \in S \cup S\Sigma^i$ the table value is set using forecasting queries:

$$T_{k+1}(s, \varepsilon) = \Phi(s, \varepsilon, w^i).$$

Check consistency

While (S, E, T) is not consistent,

Find $s_1, s_2 \in S$, $\sigma \in \Sigma^i$ and $e \in E$

Such that row(s_1) = row(s_2) and $T(s_1\sigma, e) \neq T(s_2\sigma, e)$,

$E_{k+1} = E_k \cup \{\sigma e\}$

$\forall s \in S \cup S\Sigma^i$ the table value is set using forecasting queries:

$$T_{k+1}(s, \sigma e) = \Phi(s, \sigma e, w^i).$$

Check completeness

While (S, E, T) is not complete

Find $s \in S\Sigma^i$ such that for all $s' \in S$ $row(s') \neq row(s)$:

$S_{k+1} = S_k \cup \{s\}$

$\forall\sigma \in \Sigma_k^i$, $S\Sigma_{k+1}^i = S\Sigma_k^i \cup \{s\sigma\}$

The table value is set using forecasting queries

$\forall e \in E$, $T_{k+1}(s\sigma, e) = \Phi(s\sigma, e, w^i)$.

Return $P^i(S, E, T)$

$\Phi(s, e, w^i)$:- returns the expected outcome of the example $\phi^i(s, e)$

Given a stream of data received during the interactions with the environment, D^i, a player builds a table T that summarises the perceived residual product automaton, P^i. This table contains the whole set of data, both the observed stream of actions and their outputs, and the forecasted actions with the expected outcomes. In table T, S represents the stream of actions played by player i (and the table contains the received outcomes) while $S\Sigma_k^i$ represents the forecasted stream of actions (and the table contains the expected outcomes).

It is noteworthy that a player is able to observe the outcomes of the actions along the played path and to infer the outcome of his actions off-equilibrium path, with a certain error probability, w^i. This w^i corresponds to a "trembling perception" i.e., the player can forecast the outcomes of the off-equilibrium moves but, with a certain probability, he forecasts the wrong outcome. In other words, the forecasting function $\Phi(s, e, w^i)$ returns the true outcome if the other players keep the same automaton and if there is no forecasting tremble.

The algorithm ends by checking table T for completeness and consistency. The algorithm terminates with the inference of a complete and consistent residual product automaton.

In summary, a QPR player is able to collect all the information needed to build a complete and consistent model of his opponent(s). First, he updates the automaton model using information gathered in the interaction with the environment. Second, he infers the possible outcomes associated with each possible move at every state of the automaton model. Since the QPR algorithm can model games where only a small number of iterations with the environment are available, it represents a very powerful tool to simulate learning in games. This is not the case of reinforcement learning which, in order to be successful, needs to try actions off-equilibrium path. While a QPR player can infer the off-equilibrium outcomes, it may take several iterations to learn them with an algorithm of reinforcement learning and, most importantly, the QPR player avoids the trial-and-error learning that implies experimenting actions that may be irreversible or very costly.

Moreover, it is noteworthy that independent of the nature of the residual product automaton (stochastic or deterministic), a player can use the QPR identification algorithm as a learning tool. This is important as in economic problems *nature* is usually one of the agents which needs to be taken into account: an agent of the type *nature* is usually characterised by exhibiting random behaviour. In this case, in order to describe nature's behaviour, a learning algorithm infers a probabilistic residual product automaton, specifically a probabilistic output automaton (Ron and Rubinfeld, 1997).

TABLE 6.3: The true residual product automaton M

		δ^m	
Q^m	λ^m	a	b
q_1^m	0	q_1^m	q_2^m
q_2^m	0	q_3^m	q_4^m
q_3^m	0	q_5^m	q_4^m
q_4^m	1	q_2^m	q_3^m
q_5^m	1	q_1^m	q_5^m

Next, an example illustrates the QPR identification algorithm. Assume an environment with deterministic outcomes where an agent has two possible actions $\Sigma^i = \{a, b\}$, and where the environment has only two possible outcomes, $\Sigma^m = \{0,1\}$. Finally, assume that $w^i = 0$ for every *i*. Table 6.3 describes the *true* residual product automaton $M = \left(Q^m, q_0^m, \Sigma^m, \delta^m, \lambda^m\right)$.

Table 6.4 represents the agent *i*'s initial model of the residual product automaton $M_1^i = \left(Q^{mi}, q_0^{mi}, \Sigma^{mi}, \delta^{mi}, \lambda^{mi}\right)$.

TABLE 6.4: M_1^i

		δ^{mi}	
Q^{mi}	λ^{mi}	a	b
q_1^{mi}	0	q_1^{mi}	q_1^{mi}

This model is compatible with the data available: $D^i = \{(\varepsilon,0),(a,0),(ab,0)\}$. This player then plays *b*, and therefore defines a string *abb* and gets a counterexample $\phi^i = (abb,1)$, which makes him revise the inferred model. The inference process starts by constructing the table (*S, E, T*) as shown in Table 6.5 (it includes an extra column Q^{mi} that classifies each row into the corresponding state of the model, M^i).

Then this player proceeds following the algorithm in Table 6.2. First step, he looks for changes in the residual product automaton: M^i did not change. Then he checks the consistency and closing of the new model. The model in Table 6.5 is not consistent since $row(\varepsilon) = row(ab)$ and $row(b) \neq row(abb)$, i.e., the same action *b*, applied in the same state, q_1^{mi}, leads to different outcomes. In order to solve the inconsistency this player adds another test, action *b*, to the tests set, *E*.

TABLE 6.5: (*S, E, T*) revised table

		E	Q^{mi}
		ε	
S	ε	0	q_1^{mi}
	a	0	q_1^{mi}
	ab	0	q_1^{mi}
	abb	1	q_2^{mi}
SΣ	a	0	q_1^{mi}
	b	0	q_1^{mi}
	aa	0	q_1^{mi}
	aba	0	q_1^{mi}
	abba	0	q_1^{mi}
	abbb	0	q_1^{mi}

Hence, Table 6.6 represents the new model (*S, E, T*) complete and consistent with the new data. One important feature of this new model is that it has three states while the initial model only has one.

TABLE 6.6: (*S*, *E*, *T*) table, including an extra test to correct inconsistencies

		E		Q^{mi}
		ε	b	
S	ε	0	0	q_1^{mi}
	a	0	0	q_1^{mi}
	ab	0	1	q_2^{mi}
	abb	1	0	q_3^{mi}
$S\Sigma$	a	0	0	q_1^{mi}
	b	0	1	q_2^{mi}
	aa	0	0	q_1^{mi}
	aba	0	1	q_2^{mi}
	abba	0	1	q_2^{mi}
	abbb	0	1	q_2^{mi}

Table 6.7 represents the new residual product automaton $M_2^i = \left(Q^{mi}, q_0^{mi}, \Sigma^{mi}, \delta^{mi}, \lambda^{mi}\right)$, where $q_0^{mi} = q_3^{mi}$. However, even assuming that the players have no perception biases, i.e., $w^i = 0$ for all i, a player can still commit inference errors due to his limited experience: the inferred complete and consistent algorithm is not the true one.

TABLE 6.7: Residual product automaton M_2^i

Q^{mi}	λ^{mi}	δ^{mi} a	δ^{mi} b
q_1^{mi}	0	q_1^{mi}	q_2^{mi}
q_2^{mi}	0	q_2^{mi}	q_3^{mi}
q_3^{mi}	1	q_2^{mi}	q_2^{mi}

Thus, it is noteworthy that an agent i, although rational, (he builds a complete and consistent model) cannot infer the correct model of the environment given his lack of experience. In addition, this agent has no guarantee of learning the *true* residual product automaton, M, since he is not provided with specifically chosen counter-examples that would allow him to do so (his only sources of information are the outcomes of interacting with other players).

6.3. The Adaptation Algorithm

This section presents the adaptive best response (ABR) algorithm, starting by defining the concepts of plausible model (Definition 6.3) and passive inertia (Definition 6.4).

Suppose that a player i always uses the same QPR algorithm to infer a complete and consistent model (P^i) of the environment. Let, for a given player i,

$\boldsymbol{P}_t^i = \{P_1^i, P_2^i, ..., P_t^i\}$, $\boldsymbol{D}_t^i = \{D_1^i, D_2^i, ..., D_t^i\}$ and $\boldsymbol{A}_t^i = \{A_1^i, A_2^i, ..., A_t^i\}$ represent, respectively, the set of past models inferred, the set of outputs received, and the set of automata used during the game. Further, let Φ represent a plausibility function that classifies the goodness of the QPR algorithm as an instrument to forecast the future behaviour of the system. Thus $\Phi: \boldsymbol{P}_t^i \times \boldsymbol{D}_t^i \times \boldsymbol{A}_t^i \rightarrow \{0,1\}$ classifies the QPR algorithm, at a certain time *t*, as a reliable instrument (1) or as an unreliable instrument (0).

DEFINITION 6.3 (Plausible Model and Plausible configuration): *Therefore, a model* P_t^i *is said plausible if and only if* $\Phi\left(\boldsymbol{P}_{t-1}^i, \boldsymbol{D}_{t-1}^i, \boldsymbol{A}_{t-1}^i\right) = 1$, and not plausible otherwise. *A state of the automata game, i.e., one of the possible configurations of W, is plausible when every player i holds a plausible model of the environment.*

Therefore, a model P_t^i is plausible if a QPR-learner perceives it to be a good representation of M_t^i. Definition 6.3 emphasises that the plausibility of a certain model depends on the credibility of the inference mechanism, i.e., on the past successes of the models inferred by the QPR algorithm. Next, Proposition 6.2 proves that a justification for inertia is non-plausibility of a certain model (inferred by using a QPR algorithm). Further, Proposition 6.3 shows that inertia is a necessary condition for rational behaviour when uncertainty a model is not plausible.

DEFINITION 6.4 (Passive Inertia): *Passive inertia is the behaviour of a player i that does not follow the model suggested by best response, i.e.,* $A_{t,BR}^i = BR\left(P_t^i\right)$, *as the* *inferred model,* P_t^i*, is not plausible. Instead, this player i keeps the same automaton, i.e.,* $A_t^i = A_{t-1}^i$.

Thus, passive inertia is related to Eliaz's (2003) criticism of the Nash equilibrium (which assumes that players have correct forecasts). Hence, passive inertia represents the attitude of a rational agent that does not have enough evidence to support a change in his behaviour and needs to collect more data before adapting to the environment.

PROPOSITION 6.2: *Assuming the existence of an infinitesimal cost of changing automaton, when a model* P_t^i *is not plausible, passive inertia is a necessary condition for rational behaviour.*

PROOF: Let $A_{t,BR}^i$ stand for the best-response automaton against the perceived model P_t^i, i.e., $A_{t,BR}^i \equiv BR(P_t^i)$. Further, assume that the model P_t^i is not plausible. In this case, a preference on the outcome of best response or passive inertia cannot be imposed as, for this player, $\sum_{j=t}^{+\infty} \rho_i^{j-t} E_t\left(u_j^i\left(z_j^i \setminus A_{t,BR}^i\right)\right) \neq \sum_{j=t}^{+\infty} \rho_i^{j-t} E_t\left(u_j^i\left(z_j^i \setminus BR\left(M_t^j\right)\right)\right)$.

Hence, due to the infinitesimal cost of change, and by definition of rational behaviour, the player would keep the same automaton, i.e., $A_t^i = A_{t-1}^i$. *Q.E.D.*

Another facet of this uncertainty (which is generated by boundedly rational behaviour) is the possibility of a player to use active inertia to signal the credibility of his own behaviour[31]. Active inertia, Definition 6.5, implies that keeping a stable behaviour increases a player's credibility, while best response behaviour when leading to frequent automaton switching decreases a player's credibility.

DEFINITION 6.5 (active inertia): *Active inertia is the behaviour of a player i that keeps the same automaton, i.e.,* $A_t^i = A_{t-1}^i$, *not adapting by best response, i.e.,* $A_{t,BR}^i = BR\left(P_t^i\right)$, *even if* P_t^i *is plausible.*

A player uses active inertia when there are multi-equilibria and when he wants to restrict the solutions available to his opponents. A player uses active inertia to convey credibility and, in this manner, aiming to impose his behaviour to other players.

Let $z_j^i \setminus BR(P_t^i)$ and $\zeta_j^i \setminus A_t^i$ represent the outcome received by player i at iteration j, respectively by following best-response behaviour to his perceptions at iteration t, and by keeping the same automaton (and allowing the other players to adapt to his behaviour by using best-response).

[31] Active inertia is preferred to Kalai and Leher's (1993) "teaching behaviour" in order to avoid a possible confusion with the teacher in the "automata learning" theory (Angluin, 1987). It is also a similar concept to Gilboa and Samet's (1989) tyranny of the weak.

PROPOSITION 6.3: *Active inertia is a necessary condition for player i's rational behaviour if the expected payoff of playing best-response is not higher than the expected payoff of keeping the same automaton, i.e.,*

$$\sum_{j=t}^{+\infty} \rho_i^{j-t} E_t\left(u_j^i\left(z_j^i \setminus BR(P_t^i)\right)\right) \leq \sum_{j=t}^{+\infty} \rho_i^{j-t} E_t\left(u_j^i\left(\zeta_j^i \setminus A_t^i\right)\right).$$

PROOF: Let $A_{t,BR}^i$ stand for the best-response automaton against the perceived model P_t^i, i.e., $A_{t,BR}^i \equiv BR\left(P_t^i\right)$. Further, assume that $A_{t,BR}^i \neq A_t^i$, and P_t^i is plausible, thus player i expects $\sum_{j=t}^{+\infty} \rho_i^{j-t} E_t\left(u_j^i\left(z_j^i \setminus A_{t,BR}^i\right)\right) > \sum_{j=t}^{+\infty} \rho_i^{j-t} E_t\left(u_j^i\left(z_j^i \setminus A_t^i\right)\right)$. However, by keeping A_t^i, player i increases the plausibility of his behaviour, as for any player $z \neq i$ the residual product automaton M_t^z depends on the automata used by other players. If these automata are stable then M_t^z is stable as well, and P_t^z is plausible. Therefore, further assuming that player z does not use active inertia, he plays $A_{t,BR}^z = BR\left(P_t^z\right)$, adapting his behaviour to A_{t-1}^i, and therefore changing the outcomes received by i from z_t^i to ζ_t^i, hence:

$$\sum_{j=t}^{+\infty} \rho_i^{j-t} E_t\left(u_j^i\left(z_j^i \setminus BR(P_t^i)\right)\right) \leq \sum_{j=t}^{+\infty} \rho_i^{j-t} E_t\left(u_j^i\left(\zeta_j^i \setminus A_t^i\right)\right). \; Q.E.D.$$

Following the automata identification literature, the construction of the ABR algorithm obeys three principles: Markov property, Compactness and Completeness. The Markov property imposes that the action, at each possible state of the automaton, is independent of past actions and past internal states of the automaton. Compactness reflects the Occam's razor principle: the rule describing a player's behaviour must be as simple as possible. Completeness prescribes that an automaton needs to exhaustively define actions for every possible state. Furthermore, the ABR algorithm applies the principles of passive and active inertia, and the principle of best response in order to model rational behaviour.

THEOREM 6.1: *Best response is not a sufficient condition for rational behaviour.*

PROOF: It follows from Propositions 6.1 and 6.2.

An FADG player faces an infinite stage optimisation problem with discounting (where ρ_i represents player *i's* discount factor). Banks and Sundaram (1990), and Piccione (1992) model the infinite stage game with discounting using dynamic programming. The decision variables are the actions of player *i* in each state of the product automaton, $\lambda^i : Q^{mi} \to \Sigma^i$. The player *i'* s objective is to compute the automaton A^i that generates a policy with the highest discounted reward at time *t*, V_t^i, subject to the constraints imposed by P^i, i.e., $\max \sum_{t=0}^{+\infty} \rho_i^t u^i\left(\lambda^{pi}\left(q_t^{pi}\right)\right)$ subject to $q_{t+1}^{pi} = \delta^{pi}\left(q_t^{pi}, a_t^i\right)$ where a_t^i represents player *i*'s action at time *t*, and q_0^{pi} is the initial state of P^i. The function V_t^i may be formalised recursively using dynamic programming. The player maximises $V_1^i\left(q_0^{pi}\right)$ where $V_t^i\left(q_t^{pi}\right) = u_t^i\left(\lambda^{pi}\left(q_t^{pi}\right)\right) + \rho_i^t \max_{a_t^i} V_{t+1}^i\left(\delta^{pi}\left(q_t^{pi}, a_t^i\right)\right)$ for *t* =*1*, …, +∞ subject to $q_{t+1}^{pi} = \delta^{pi}\left(q_t^{pi}, a_t^i\right)$. Table 6.8 presents these equations within the Adaptive Best Response algorithm.

TABLE 6.8: Adaptive Best Response algorithm

ABR(A^i, P^i, V^i, ρ_i) algorithm

A^i: The current player i's automaton, $A^i = \left(Q^i, q_0^i, \Sigma^i, \delta^i, \lambda^i\right)$

P^i: The perceived residual product automaton

ρ_i: Discount factor for player i, $0 \le \rho_i \le 1$

S^i: Optimal policies generated by the automaton A^i

V_t^i: Discounted reward at time t,

Each player i evaluates the plausibility of P^i

If P^i is plausible and if the player is not using Active Inertia

$A^i = BR(P^i)$

Otherwise: $A_t^i = A_{t-1}^i$

Algorithm Best-Response $A^i = BR(P^i)$:

Compute S^i the optimal policy play against P^i:

$$S^i = \arg\max_{a_t^i}\left[u_t^i\left(\lambda^{pi}\left(q_t^{pi}\right)\right) + \rho_i V_{t+1}^i\left(\delta^{pi}\left(q_t^{pi}, a_t^i\right)\right)\right]$$

$s.t.$

$$q_{t+1}^{pi} = \delta^{pi}\left(q_t^{pi}, a_t^i\right)$$

$$q_1^{pi} = q_0^{pi}$$

Compute A^i from the optimal policy S^i

Let g represent the optimal policy such that

$$g : Q^{pi} \to S^i$$

$$S^i = g\left(Q^{pi}\right)$$

The automaton A^i has the same number of states as M^i

$$Q^i = Q^{pi} \Leftrightarrow q_t^i = q_t^{pi}$$

Define the initial state and action of A^i

$$q_0^i = q_0^{pi}$$

Define the behaviour function: assign an action to each state of A^i

$$\lambda^i\left(q_t^i\right) = g\left(q_t^{pi}\right)$$

Define the transition function

$$\delta^i\left(q_t^i, \lambda^{pi}\left(q_t^{pi}\right)\right) = \delta^{pi}\left(q_t^{pi}, \lambda^i\left(q_t^{pi}\right)\right)$$

The ABR algorithm starts by applying the principles of active and passive inertia. A player only follows the best response principle if the environment is stable (in order for the best response to be meaningful), and if he did not change his behaviour recently (in order to allow other players to adapt to his new automaton), this is active inertia behaviour. If the environment is stable, and if other players have had enough time to adapt, a player computes the best response to the perceived model of the environment, building a new adaptation automaton, A^i.

The best response algorithm associates each state and transition in P^i to a new state and transition in A^i, constructing a one-to-one correspondence (Banks and Sundaram, 1990).

Next, an example illustrates ABR algorithm. Assume a player i with a discount rate $\rho^i = 0.9$ holding the model of the residual product automaton M_2^i (Table 6.7). Assume that $u^i\left(z^m\right) = z^m$, i.e., the utility of a certain outcome is that outcome. Then, Table 6.9 depicts the automaton that maximises an agent i's discounted reward, i.e., $A_1^i = (Q^i, q_0^i, \Sigma^i, \delta^i, \lambda^i)$, which is computed using the adaptation algorithm, and where $q_0^i = q_3^i$.

TABLE 6.9: Automaton A_1^i

		δ^i	
Q^i	λ^i	*0*	*1*
q_3^i	*a, b*		q_2^i
q_2^i	*b*	q_3^i	

However, the automaton A_1^i is not compact and generates more than one policy, since it defines in state q_3^i it may take different actions. Agent i's next task is to make the automaton A_1^i compact. The algorithm reduces the number of state in this automaton by choosing action b in state q_3^i. Table 6.10 presents the new automaton $A_2^i = (Q^i, q_0^i, \Sigma^i, \delta^i, \lambda^i)$, which is compact and where $q_0^i = q_3^i$.

TABLE 6.10: Automaton A_2^i

		δ^i	
Q^i	λ^i	*0*	*1*
q_3^i	b	q_3^i	q_3^i

Finally, the automaton A_2^i, although more compact, is not deterministic anymore. The same action, in the same state, has different outcomes (this is a probabilistic output automaton).

6.4 Analysis of the FADG Dynamics

This section analyses the convergence properties of the FADG. Theorem 6.2 shows that the combination of best response dynamics with passive and active inertia implies that the players exhibit rational behaviour throughout the game trajectory. Further, the properties of the game trajectory are analysed, as well as the impact of inference forecasting errors on the stability of equilibrium solutions.

THEOREM 6.2: *In the trajectory of an FADG every player exhibits rational behaviour.*

PROOF: During the FADG described in Table 6.1, a player i can choose between three possible types of behaviour, depending on P^i's plausibility. When the inferred model is non-plausible, he keeps $A_t^i = A_{t-1}^i$, which by Proposition 6.2 is the rational choice. When the inferred model is plausible, a player i has two possible choices: the best response behaviour $A_{t,BR}^i = BR\left(P_t^i\right)$, which is rational by definition of best response, or to follow an active inertia strategy, which is rational behaviour by Proposition 6.3. *Q.E.D.*

The trajectory of the game and the equilibrium properties depend on the strength of the inertia principles (passive and active). When passive inertia is strong, the trajectory tends to present transient stationary states, in which the length of the stationary period is a function of the inertia strength and the forecasting errors in the QPR algorithm. On the other hand, active inertia shapes the game trajectory by acting together with best-response behaviour. The latter changes the product automaton W by modifying a player's automaton. Then, if after a player changing his automaton a game is not in equilibrium, each one of the other players can reply by changing his automaton as well. As a result a new transitory stationary state appears.

However, if best response and inertia are the main drivers of the behaviour of a game, it is also true that the "perception trembles" can influence the behaviour of the game. Theorem 6.3 addresses this issue. Let w^i stand for the forecasting error probability of a given player i, i.e., the probability that he commits an error by forecasting the wrong outcome of an action.

THEOREM 6.3: *The higher the value of* w^i*, the higher the instability of* W *will be, in equilibrium.*

PROOF: A higher forecasting error (w^i) implies that the inference of a model P^i probably contains an increased number of mistakes regarding the forecasts of off-equilibrium path outcomes. Thus, best response, i.e., $A^i_{t,BR} = BR\left(P^i_t\right)$, leads to sub-optimal behaviour, as $P^i_t \neq M^i_t$ with a higher probability. Moreover, assume that a certain configuration W is a conjectural equilibrium, and that a player i infers a model P^i of the environment. If due to the forecasting error w^i it happens that $A^i_{t+1,BR}(P^i_{t+1}) \neq A^i_{t,BR}(P^i_t)$, then the system's configuration (W) changes, even if the system was in a conjectural equilibrium. Hence, increased probabilities of forecasting errors lead to unstable equilibria. *Q.E.D.*

Finally, depending on the specificities of the problem modelled, a game may eventually converge to a Nash or conjectural equilibrium. However, even if the game eventually converges to equilibrium, the trajectory may be long and full of transient stationary configurations. Hence, the trajectory path can be more important than the equilibrium itself.

6.5 Simulation of Two Games Using the FADG

This section simulates the Keynesian coordination game and the pie game using the FADG framework. Let α_a^i and α_v^i represent the smoothing rate for the average present value and for its volatility, respectively. In these simulations, passive inertia is modelled using the smoothed average of the automaton's present value, i.e., $m_t^i = \alpha_a^i m_{t-1}^i + (1-\alpha_a^i)V_t^i$, and its volatility, i.e., $v_t^i = \alpha_v^i v_{t-1}^i + (1-\alpha_v^i)\left|V_t^i - m_t^i\right|$, thus capturing the instability of the player's forecasts. The volatility of the present value represents a proxy for the forecasting errors, capturing the instability of other players' behaviour and the plausibility of a certain P^i.

Let θ^i represent a parameter such that $0<\theta^i<1$ and I_t^i represent the probability of using active inertia:

$$I_t^i = \begin{cases} 1 & \Leftarrow A_t^i \neq A_{t-1}^i \\ \theta^i I_{t-1}^i & \Leftarrow A_t^i = A_{t-1}^i \end{cases}.$$

This means that a player, after changing his behaviour, waits for the other players to adapt to his new automaton. If the environment is stable and I_t^i is low, the player changes his behaviour using the best-response algorithm in Table 6.8.

van Huyck et al. (1990) and Cooper et al. (1990) analyse the problem of coordination in strictly Pareto ranked games using an experimental environment. They report that in their experiments the players could not coordinate their behaviour and that the game converges to the equilibrium with the most inefficient outcome, i.e., the security-payoff proposed by von Neumann and Morgenstern (1953), instead of converging to the payoff-dominant equilibrium proposed by Harsanyi and Selten (1988).

The simulations of this Keynesian game included 16 players. The utility (u^i) of each player i is $u^i = 0.6 + 0.2\min(B) - b^i$, where B represents the set of all bids and b^i stands for player i's bid. The parameters were defined as $\Sigma^i = \{1,...,7\}$, $\rho_i = \theta^i = 0.9$, $\alpha_a^i = \alpha_v^i = 0.25$, with the initial bids between three and seven. Figure 6.1 displays the evolution of the average bid in three games. These experiments assumed that the forecasting error probabilities (w^i) are, respectively, zero, one and five percent.

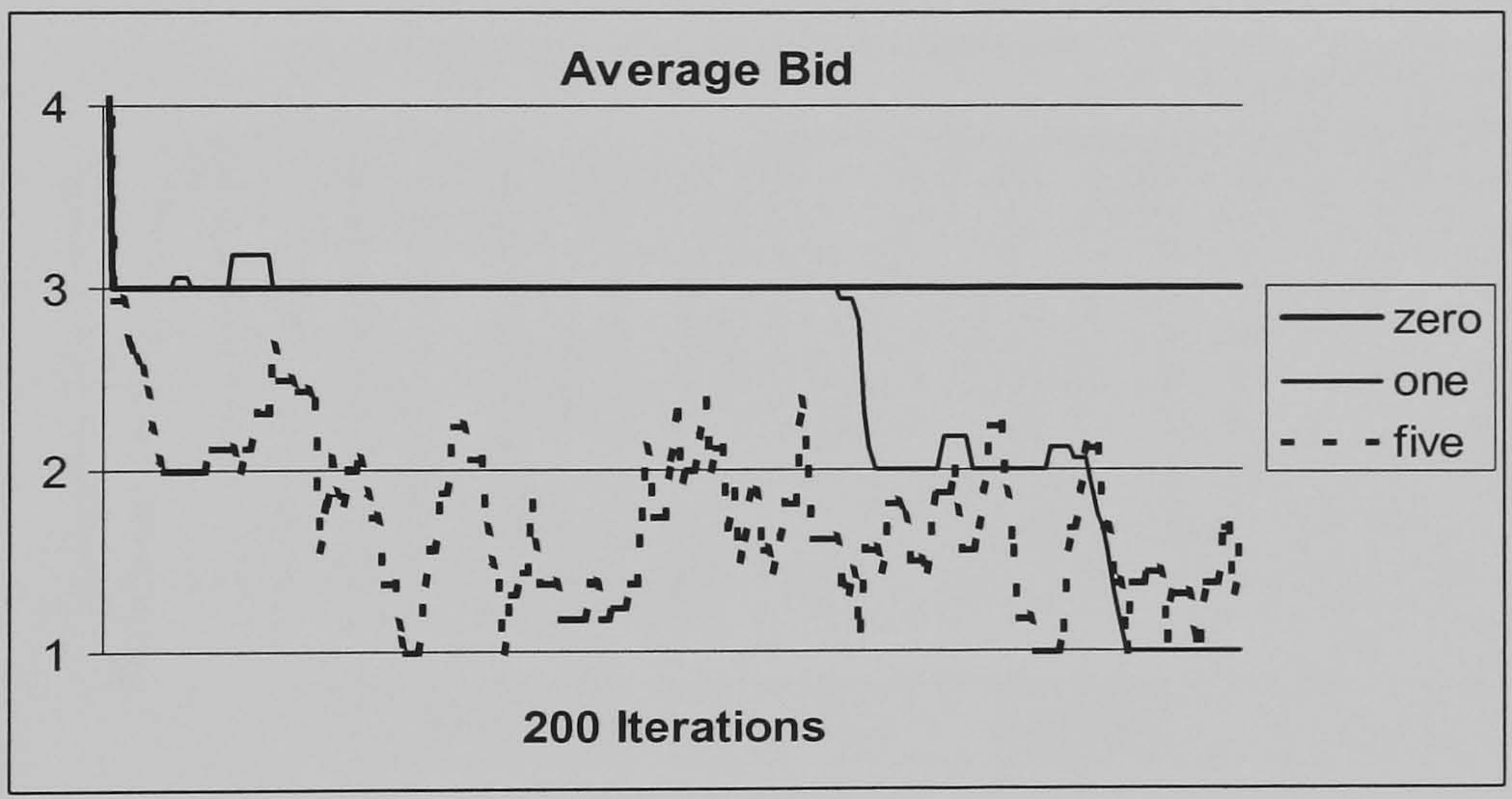


FIGURE 6.1: Keynesian Coordination Game

The experiment replicated in Figure 6.1 corresponds to van Huyck et al.'s (1990: 240) treatment A. In their experiments, after 40 simulations of the same game under the same treatment, the average converged to 1.2 and the minimum bid was one. The results with the five percent forecasting error are the ones that more closely replicate the results of the experiments with people.

The pie game, Figure 6.2, was simulated with six players. The utility $\left(u^i\right)$ of each player i is

$$u^i = \begin{cases} b^i & \leftarrow \sum_{b^i \in B} b^i \le 15 \\ -b^i & \leftarrow Otherwise \end{cases},$$

and the parameters were defined as $\Sigma^i = \{1,2,3\}$, $\rho_i = \theta^i = 0.9$, $\alpha_a^i = \alpha_v^i = 0.25$, and the pie had 15 pieces in total.

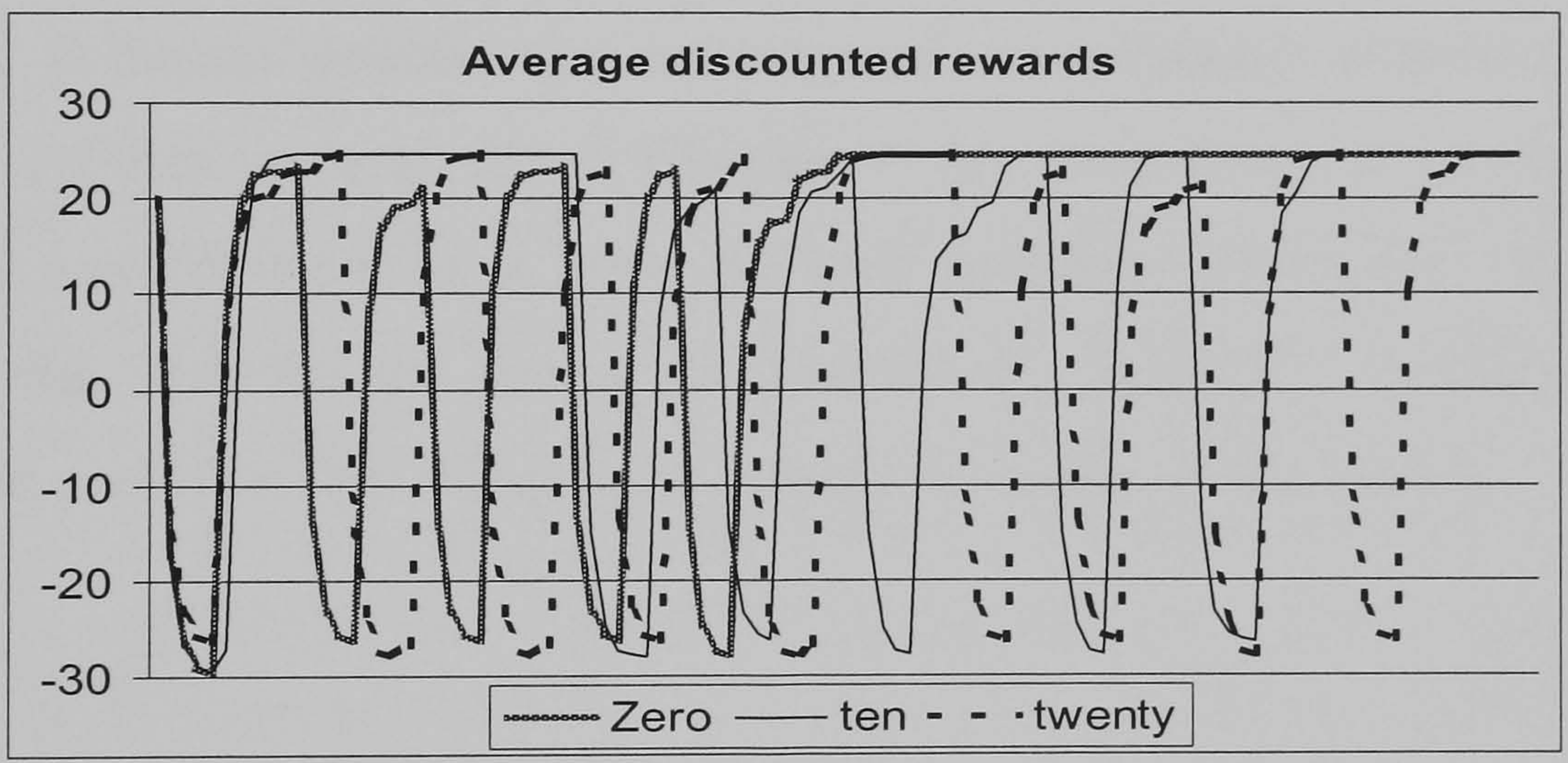


FIGURE 6.2: The Pie game

The game where players exhibit no forecasting error converged to the Pareto-Nash equilibrium, while the game where players exhibit forecasting errors fluctuates around the equilibrium (see Figure 6.2). Furthermore, the lower the forecasting error the more stable is the equilibrium.

So, why did the pie game converge to the Pareto-Nash equilibrium, and why did the Keynesian coordination game fail to converge to the payoff-dominant equilibrium? The answer lies in the structure of these two games. While in the Keynesian coordination game, coordination is very demanding, as one player is enough to ruin the "collusive outcome", in the pie game several players are needed for the Pareto-Nash equilibrium not to hold.

Therefore, the experiments show that boundedly rational players cannot coordinate their behaviour in games where the defection of a small number of players changes the equilibrium. Hence, the results in both games support QPR as a good approximation of boundedly rational behaviour: the game converges to a Nash equilibrium when there is no forecasting error, and it fails to converge, otherwise.

Finally, it should be noticed the importance of coordination games in electricity economics. For example, this is the case in Chapters 4 and 5, where the occurrence of implicit collusion implies the existence of coordination without an explicit agreement. Further, as analysed in Chapter 7, the plant trade, is by itself implies behavioural coordination in a very difficult problem. Moreover, we can also speculate that investment games in liberalised electricity markets are also coordination problems (but this issue is outside the scope of this thesis).

6.6. Conclusion

Some literature in automata games tends to assume the existence of an exogenous automata set from which a player can choose his strategy. Some other literature assumes a player that computes the best-response strategy against a given automaton used by his opponent. However, in the first case, it is never explained why the players have common knowledge of a given automata set. On the other hand, in the second case, it is not explained why the optimising player knows that his opponent is an automaton.

This chapter presents a new framework to model learning and adaptation in N-player extensive form games of incomplete information, the Finite Automata Dynamic Game (FADG). An FADG models the set of possible automata as endogenous variables, capturing the process through which a model of other players' behaviour is learned, and enabling the analysis of the process from which certain types of behaviour emerge. The players' learning and adaptation capabilities determine the possible automata used by them.

Each player, in order to infer a model of the environment (the perceived residual product automaton), uses a Quasi-Perfect-rationality (QPR) algorithm, and has the ability to redefine his behaviour by changing his automaton (using the Adaptive Best-Response algorithm). Therefore, the FADG enables the modelling of co-evolutionary automata that learn and adapt together, reshaping the product automaton of the environment. Hence, the FADG provides the missing link in the automata games literature: the bridge between inference of the opponents' automata and learning behaviour.

The main contributions of the FADG as a modelling technique are as follows. A) The product automaton defining the game is endogenous, being a function of the players' behaviour. B) The automaton used by each player changes over time, which means that the possible internal states, the transition function, and the behavioural functions are co-evolving with the system. C) There is no common knowledge of the true set of states of the system. A player adapts to his perceptions and defines endogenously the internal set of states of his automaton, and the transition and behavioural functions. D) Given the data received from a given number of interactions, a player is able to infer a model of the system that is not restricted to the set of outcomes and actions played in the history of the game. Given the QPR algorithm, a player is able to infer how the environment would evolve for actions off-equilibrium that may be costly to take. E) Another contribution of this paper is the concept of residual product automaton. This allows a player to infer a model of the system as a whole instead of modelling each one of his competitors separately. Further, it enables learning and adaptation in games with a large number of players. F) Through the QPR algorithm, each agent may infer the value of a certain action without actually experimenting it. This is a major point of this learning algorithm, as by using reinforcement learning a player needs to try actions that a person or organisation would not attempt. G) Finally, the FADG captures the uncertainty generated by boundedly rational behaviour.

At a behavioural level, Proposition 6.1 shows that a necessary condition for rational behaviour is the capability to infer a complete and consistent model of the environment and Theorem 6.1 shows that best response is not a sufficient condition for rational behaviour. The stability properties of the product automaton are also analysed and Theorem 6.3 shows that the inference forecasting errors increase the instability of the product automaton. Overall, the analysis of the FADG shows that the properties of the system as a whole, i.e., the product automaton, are a function of the learning and adaptation capabilities of the players in it.

Finally, this chapter studies learning and adaptation in two coordination games where the players are learning automata.

Hence, by inferring a model of the system's behaviour and by computing the value of playing against that model (by using a dynamic programming or a reinforcement learning algorithm), the FADG player can, for example, play the plant trading game presented in Chapter 7.

CHAPTER 7

STRAGEGIC MANAGEMENT OF PLANT PORTFOLIOS AND MARKET STRUCTURE EVOLUTION

The recent evolution of the England and Wales (E&W) and California markets (e.g., Ishii and Yan, 2002, September) suggests that companies are active in trading electricity plants among themselves in their quest for the "optimal portfolio." If this plant trading activity is associated with divestments that are expected to conduce to increased competition (e.g., Green and Newbery, 1992; Borenstein and Bushnell, 1999; Day and Bunn, 2001), it has also been identified as a possible threat to investment, particularly, Ishii and Yan present evidence that divestment has *crowded out* new investment in the California market.

This chapter addresses the issue of strategic management of plant portfolios as a motivation for plant trading and analyses its relation with market structure evolution in the short-run. The model here presented incorporates two main components: a plant trading game and an electricity market game. The plant trading game simulates the interaction between electricity companies that trade generation plants. The electricity market game simulates an electricity market assuming Cournot players.

The contributions of this paper are both methodological and theoretical.

At a methodological level, by developing a plant trading game this paper presents the following contributions. First, it simulates how players learn and adapt in order to trade electricity plants. This is crucial to modelling of plant trading, as the actions in this game are very costly and the players rarely interact. Second, it splits the adaptation from the learning (identification) processes. Third, it models the evolution of market structure in off-equilibrium states of the industry, as this is a requirement for plant trading to occur. Fourth, it captures the attitude of players attempting to coordinate their behaviour in a game with a very large number of possible actions.

As, by definition, in a trading game all trading takes place in off-equilibrium states of the industry, the trajectory toward equilibrium is as important as the equilibrium itself. Thus, in order to explain plant trading one needs to look at dynamic models that incorporate past dependencies, learning and adaptation. Hence, the theory of learning in games seems to be the tool to use to model plant trading, as it aims to explain how equilibrium arises as an outcome of the interactions between boundedly rational players. A second motivation to use learning theory in this setting is that the plant trading game is computationally hard (see Proposition 7.5).

Moreover, there are several other arguments that have been put forward as reasons to look at models of learning (e.g., Roth and Erev, 1995; Samuelson, 1997: 15; Fudenberg and Levine, 1998: 1; Feltovich, 2000; Sarin and Vahid, 2001). First, people do not always behave rationally: strategies emerge as a trial and error process in which players find that some strategies behave better than other strategies do. Second, equilibrium theory does a poor job in explaining how agents play in the early stages of most games, although it explains much better the behaviour in later stages.

However, due to the high number of iterations required by models of learning these cannot handle properly games where players seldom interact (Samuelson, 1997; Fudenberg and Levine, 1998). For this reason, in order to model the learning process in the plant trading game, this chapter follows the finite automata game presented in Chapter 6 as it allows the development of models of learning in a setting where players rarely interact and (or) actions are potentially irreversible.

At a theoretical level, the contributions of this paper are threefold.

First, this paper develops a theoretical model for analysing the economic value of a plant. This analysis extends the current theory on capacity withholding by studying its relation with plant trading. Specifically, the concept of economic value of a plant (i.e., *operational profit* and *portfolio contribution*) extends Joskow and Kahn's (2002) analysis (which is focused on the supply function slope) by decomposing the value function into finer components. Moreover, the model developed also examines how different types of plants contribute differently to the value of the portfolio as a whole. Further, it also explains how a plant has different value for different players.

Second, it analyses how the plant trading dynamics depends on the economic value of plants. Particularly, by examining how the value of a plant depends on the type of portfolio, it gives insights into the use of plant trading as a way to achieve a dominant position. Specifically, it shows that if trading of plant increases market efficiency (in the sense of calling first the most efficient plants) it also implies increased capacity withholding.

Third, it develops a theoretical analysis of the impact of market design on the evolution of market structure. Specifically, it analyses at the main attractors of a dynamic system under single-clearing and multi-clearing mechanisms. This comparison looks at the issue of discriminatory vs. uniform pricing. However, this model is developed on the hypothesis that price discrimination is due to technological constraints and it does not arise from different types of auction (as assumed in Bower and Bunn, 2000; Abbink et al., 2003; Rassenti et al., 2003). Instead, the model in this paper follows Borenstein et al. (1995) who suggested that in the liberalised power markets there would be different markets for baseload and peak plants, and further it also follows Elmaghraby and Oren (1999) who have proposed a clearing mechanism that implies discriminatory pricing by technology.

Next, section 7.1 presents an analysis of the properties of the electricity market game. Then, section 7.2 examines the relation between plant value and capacity withholding, which is the main driver of plant trading. Section 7.3 presents the plant trading game and the different algorithms used to simulate learning, adaptation and trading. Finally, this chapter concludes with a stylised simulation of the E&W electricity market, as an illustration of the plant trading game at work.

7.1 The Electricity Market Game

The first task in developing a model of an electricity market is to ensure its validity. Are the models developed valid representations of reality? As the objective of this chapter is not to replicate any specific system, comparing the behaviour of the simulation with the behaviour of any specific market cannot determine its validity. However the question remains: why this model and not any other one?

Hence, the different models of electricity markets used in this chapter were validated by:

1. Specifying the stylised facts that we know are true in electricity markets, which are captured by the models:

    a. A generator's supply function is step-shaped.

    b. A generator may receive different prices for his generation from different plants, even if these are identical.

    c. Different generators may price the same type of plant differently.

    d. A generator aims at maximising the value of his portfolio of plants as a whole.

2. Looking at the theory in electricity markets, in which Cournot behaviour is the most commonly used. The use of the Cournot model in electricity markets is defendable at theoretical level. Kreps and Scheinkman (1983) show that the Bertrand model requires both price competition and production to take place after demand determination. Specifically, they show that the Cournot equilibrium is the outcome of games where there is a capacity pre-commitment followed by Bertrand competition. This is arguably the case of electricity markets. Further, Daughety (1985) in analysing conjectural variations shows that rational oligopoly equilibrium is, in general, a Cournot equilibrium, and *vice versa*.

3. Looking at the empirical evidence in electricity markets. There is empirical evidence from the California market (e.g., Puller, 2002, August) that firm conduct is "relatively consistent" with a Cournot pricing game.

Thus, the analysis in this chapter models the electricity market as a Cournot game where each player decides, in each possible stage of the game, how much to generate from each plant he owns. The outputs of this Cournot game are the market price and ultimately the profit of each player.

This section presents an analysis of two different Cournot games. The first is a single-clearing Cournot game in which there is a single clearing price for each hour of the day. This model simulates a game where each player defines how much to sell at each hour (for different levels of demand), given the portfolio of plant owned. The second is a multi-clearing Cournot game in which there are different clearing prices for different markets, at a certain time of the day. In the multi-clearing mechanism, each player decides how much to offer from each one of his plants in the different markets, given the durations, the different demand functions and the structure of his portfolio.

These clearing-mechanisms define a theoretical model of prices and loads in electricity markets in which the behaviour of a generator is a function of the industry structure and of his portfolio of plants.

Each player i chooses the output $\left(Q_{i,L}\right)$ in market L characterised by a certain demand (the definition of each market can be adapted to the specific needs of the issue analysed). Let $C_{i,L}$ stand for the marginal cost of player i, A_L, α_L represent the intercept and slope of the inverse demand function, and D_L stand for the duration of market L. Further, let $K_{i,L}$ stand for player i's total available capacity in market L.

Please note that in this case $C_{i,L}$ is assumed constant for a given plant, but it may be different for the different plants owned by a player. Thus, $C_{i,L}$ can be represented as a step-function, which makes the optimisation problem computationally hard. Thus, this Cournot game is a complementarity problem (Ferris and Pang, 1997; Wei and Smeers, 1999; and Hobbs, 2001).

In both models, the start-up costs and ramp rates were not explicitly taken into account[32]. Further, the annual fixed costs are not relevant for the problem in analysis, although important for the computation of the actual value of a given plant.

[32] This is a simplifying assumption that has also been used in several other studies of electricity markets (Ramos at al., 1998; Borenstein et al. 1999, 2002).

However, since these technical constraints are important to define the capability of a plant to access a given market, the model exogenously defines, for each plant, the market in which it can sell. This simplification does not change the economics of the model and decreases its complexity from a non-linear to a linear complementarity problem.

7.1.1 The Single-Clearing Cournot Game

This model assumes that there is only a clearing price at any given time of the day, and that a player receives the same price for the electricity generated by any plant selling at that time. A Cournot model of this market is quite simple. Each player receives a clearing price P_L for the quantities sold in each one of these markets. The capacity constraint, for each market, is the total capacity available for each player at that time and in market L (in the single-clearing mechanism at any time there is only one market).

Thus, for a player i, the profit (π_i) maximisation problem is represented by equations (7.1).

$$
\begin{aligned}
&\max \pi_i = \sum_L \left(P_L - C_{i,L}\right) Q_{i,L} D_L \\
&st. \\
&P_L = A_L - \alpha_L \cdot \sum_i Q_{i,L},\ \forall L \qquad\qquad (7.1) \\
&Q_{i,L} \le K_{i,L},\ \forall L \\
&Q_{i,L} \ge 0,\ \forall L
\end{aligned}
$$

Finally, it is noteworthy that the shape of the demand function and the durations define the model used. Assume that a model specifies durations in hours, if the model uses annual durations then $\sum_L D_L = 365 \times 24$, otherwise if the model is a daily one then $\sum_L D_L = 24$.

7.1.2 The Multi-Clearing Cournot Game

In the case of bilateral electricity markets, each generator has the possibility of selling the electricity of his different plants in different markets. Based on the evidence from the E&W electricity market, the baseload, shoulder, and peak plants tend to sell electricity over different timescales with different prices. The evidence seems to suggest that different technologies sell into different market segments. Thus nuclear and CCGT seem to behave as baseload plants, coal seems to behave as a shoulder plant and finally oil, OCGT and pumped storage behave as peak plants (Power UK, 2002: 20-21). Furthermore, the very high spread between high and low prices in the UKPX, and between the System Buy Price (in the Balancing Mechanism) and the UKPX prices, are indicating evidence that in the new bilateral markets flexibility has a value and that technologies may achieve different rents (Power UK, 2002: 42-44). It is also noteworthy that this evidence does not take into account the "forward market" effect where some price "discounting" may take place due to quantity trading and risk aversion.

The multi-clearing Cournot game implies that for each time of the day there are several prices. This aims to capture the different technological characteristics of different types of plants, which is one of the factors that may create different prices for the generation of different plants, even at the same hour. Hence, a player can sell the generation from any of his plants in different markets, for a given time, possibly receiving a different clearing price in each one of them. This procedure follows the model proposed by Elmaghraby and Oren (1999) and suggested by Borenstein et al. (1995), and aims to capture the interaction between different markets and technologies in defining the value of a plant.

Thus, for a player *i*, the profit (π_i) maximisation problem is represented by equations (7.2).

$$\begin{aligned}
&\max \pi_i = \sum_L \left(P_L - C_{i,L}\right) Q_{i,L} D_L \\
&st. \\
&P_L = A_L - \alpha_L . \sum_i Q_{i,L}, \ \forall L \\
&\sum_L Q_{i,L} \le K_{i,L}, \forall L \\
&Q_{i,L} \ge 0, \ \forall L
\end{aligned} \tag{7.2}$$

Thus please note that for each market L, $K_{i,L}$ stands for player *i*'s total available capacity in market *L*, given the technological constraints of each plant (which are exogenous) and the capacity affected to other markets (in the case of a plant that is able to sell in different types of market).

Again, as in the previous model, the costs are a non-linear function of production. Further the shape of the demand function and the durations of each market define the model used. Given in hours, these durations for the annual formulation are $D_1 = 365 \times 24, D2 = ...,$ $\ldots$ and if the model is a daily one the durations are $D_1 = 24, D2 = ...$ i.e., the demand is defined for the whole duration of that market, and it does not add up to the number of hours simulated, for a day or year.

Finally, it is noteworthy that in this case, for any given time, there may be several overlapping markets: these markets are created by the joint action of bilateral trading and the specificities of the different technologies. Hence, this formulation is flexible enough to accommodate different types of market structures, such as several bilateral markets, day-ahead and balancing mechanisms, and to analyse their interaction with the different type of plant selling electricity in each one of them.

7.1.3 Optimal Behaviour and Capacity Withholding

Given the optimality conditions defined by equations (7.1) and (7.2), a player's behaviour and the interactions between the two stages of the game can now be analysed. Proposition 7.1 shows that in order for a more expensive plant to be called by a Cournot player, every available cheap plant needs to be called first.

PROPOSITION 7.1: *In a Cournot game with capacity constraints, a player owning plants with different marginal costs $C_1 < .. < C_j < .. < C_P$, offers the capacity of plant j+1 only if he does not withhold capacity from plant j.*

PROOF: Assume that a plant *j+1* can offer in market *L*. Decomposing the non-linear cost components of each firm then the variable profit function of a given player is $\pi_0 = \sum_j \left(P_L - C_j\right).Q_{j,L}.D_L$. The proof follows by contradiction. Assume for all *L* $Q_{j,L} < K_{j,L}$ and $Q_{j+1,L} > 0$. If $Q_{j+1,L} \geq K_{j,L} - Q_{j,L}$ the player can improve his profit by transferring $K_{j,L} - Q_{j,L}$ load units to plant *j*; therefore $Q_{j,L} = K_{j,L}$, reaching a contradiction. Alternatively, if $Q_{j+1,L} < K_{j,L} - Q_{j,L}$ the player can improve his profit by transferring $Q_{j+1,L}$ units of load to plant *j*; thus, $Q_{j+1,L} = 0$, again reaching a contradiction. Q.E.D.

Further, Proposition 7.2 presents some conditions under which a player can profit from capacity withholding. There is empirical evidence that portfolio players have used capacity withholding in practice. Wolak and Patrick (1997, February) present statistical evidence of capacity withholding in the E&W electricity Pool. Further, Wolfram (1998) suggests that, in the E&W Pool, generators were charging higher mark-ups over their marginal units when the cheaper units of that player were ready to generate. This strategy increases the prices received by cheaper units, and represents a form of capacity withholding (Borenstein et al., 2002).

PROPOSITION 7.2: *Let* $P_L = A_L - \alpha_L.\sum_{g=1}^{P} Q_{g,L}$ *represent the inverse residual demand function of a market L. In a Cournot game, a player owning plants with different marginal costs* $C_1 <..< C_j <..< C_P$ *and not selling any electricity generated by plants j+1, j+2, ... in that market, i.e.,* $\sum_{g=j+1}^{P} Q_{g,L} = 0$, *profitably withholds capacity of plant j if and only if* $\frac{1}{\alpha_L}\left(A_L - C_{j,L}\right) < 2.K_{j,L} + \sum_{g=1}^{j-1} Q_{g,L}$.

PROOF: Decompose the operational profit into its components in different markets, and assume the available capacities for each market as given (in both models). Then player *i*'s profit is $\pi_{i,L} = \sum_{g=1}^{P}\left(P_L - C_{g,L}\right).Q_{g,L}.D_L$ and as $\sum_{g=j+1}^{P} Q_{g,L} = 0$ it follows that $\pi_{i,L} = \sum_{g=1}^{j}\left(P_L - C_{g,L}\right).Q_{g,L}.D_L$. Further, by definition of inverse residual demand $\pi_{i,L} = A_L.D_L.\sum_{g=1}^{j} Q_{g,L} - \alpha_L.D_L.\sum_{g=1}^{j}\sum_{z=1}^{j} Q_{g,L}.Q_{z,L} - D_L.\sum_{g=1}^{j} Q_{g,L}.C_{g,L}$. Therefore, by the optimality conditions $A_L.D_L - \alpha_L.D_L.\sum_{g=1}^{j-1} Q_{g,L} - 2.\alpha_L.D_L.Q_{j,L} - D_L.C_{j,L} = 0$. Hence, the optimal load for plant j is $Q_{j,L} = -\frac{1}{2}\sum_{g=1}^{j-1} Q_{g,L} + \frac{1}{2\alpha_L} A_L - \frac{1}{2\alpha_L} C_{j,L}$, which is less than $K_{j,L}$ if and only if $\frac{1}{\alpha_L}\left(A_L - C_{j,L}\right) < 2K_{j,L} + \sum_{g=1}^{j-1} Q_{g,L}$. Q.E.D.

Next, Theorem 7.1 shows that a Cournot player does not transfer demand from a cheap to an expensive plant. This implies a merit order in the use of capacity withholding for strategic management of plant portfolios.

THEOREM 7.1: *In a Cournot game, a player owning plants with different marginal costs* $C_1 < .. < C_j < ..C_g .. < C_P$ *cannot profitably transfer load from a plant j to a plant j+g, for any g, via capacity withholding.*

PROOF: From Proposition 7.2 it follows that a player offers load from plant *j+1* only if he sells the full capacity of plant *j*. Further, Proposition 7.1 implies that it is possible to withhold the capacity of plant *j* only if the player does not offer generation from plants *j+1, j+2, ..., j+P*. Hence, a player cannot profitably transfer load from cheaper to expensive plants. Q.E.D.

Further, Theorem 7.2 extends the implications of capacity withholding to the analysis of plant trading by showing that the introduction of a new plant in the portfolio has the potential to change the value of the other plants in the same portfolio.

Let $Q_{j+z,0}$ $\left(Q_{j+z,1}\right)$ and $Q_{j,0}$ $\left(Q_{j,1}\right)$ represent, respectively, the load of plants $j+z$ and j, before (after) the transaction. Besides, let P_0 (P_1) represent the electricity market price before (after) the transaction of plant i.

THEOREM 7.2: *Assume that in a Cournot game, a player owns plants with different marginal costs* $C_1 < \ldots < C_j \ldots < C_P$. *This player can profitably reduce the load of any plant* $j+z$, *for any* $z > 1$, *from a market* L *by acquiring a plant with marginal cost* $C_{j,L}$ *only if*

$$\sum_L \left(\frac{A_L - C_{j+z,L}}{\alpha_L} \right) < 2\sum_L Q_{(j+z,0),L} + \sum_L \left(Q_{(j,1),L} - Q_{(j,0),L} \right) + \sum_L \left(\sum_{\substack{g=1 \\ g \neq j}}^{j+z-1} Q_{g,L} \right).$$

PROOF: By acquiring a plant j the inverse residual demand of player i for each market L shifts and can be represent as $P_L = A_L + \alpha_L Q_{(j,0),L} - \alpha_L \sum_{g=1}^{P} Q_{g,L}$. Therefore, by definition of marginal plant $\sum_{g=j+z+1}^{P} Q_{g,L} = 0$ and $\sum_{g=1}^{j+z-1} Q_{g,L} = \sum_{\substack{g=1 \\ g \neq j}}^{j+z-1} Q_{g,L} + Q_{(j,1),L}$, and by the optimisation conditions, it follows that

$$Q_{(j+z,1),L} = -\frac{1}{2}\sum_{g=1}^{j-1} Q_{g,L} + \frac{1}{2\alpha_L}\left(A_L + \alpha_L . Q_{(j,0),L} \right) - \frac{1}{2\alpha_L} C_{j,L} \text{ and that}$$

$$Q_{(j+z,1),L} = -\frac{1}{2}\sum_{\substack{g=1 \\ g \neq j}}^{j-1} Q_{g,L} - \frac{1}{2} Q_{(j,1),L} + \frac{1}{2} Q_{(j,0),L} + \frac{1}{2\alpha_L}\left(A_L - C_{j,L} \right).$$

Then, as $Q_{(j+z,1),L} \leq Q_{(j+z,0),L}$ by adding up $Q_{j,L}$ for all L it follows that

$$\sum_L \left(\frac{A_L - C_{j+z,L}}{\alpha_L} \right) < 2\sum_L Q_{(j+z,0),L} + \sum_L \left(Q_{(j,1),L} - Q_{(j,0),L} \right) + \sum_L \left(\sum_{\substack{g=1 \\ g \neq j}}^{j+z-1} Q_{g,L} \right). \quad \text{Q.E.D.}$$

Hence, Theorem 7.2 completes Theorem 7.1 by showing that not only load transfer is not an optimal strategy for a Cournot player, but also if a player acquires (sells) cheaper technologies he drives out (drives in) the more expensive ones. Thus, Theorem 7.2 builds a bridge between the electricity market and the plant trading game. A player holding a portfolio with a large quantity generated by cheap plants tends to reduce the electricity generated by expensive plants. Section 7.2 presents a detailed analysis of this behaviour.

7.2 Plant Value and Market Structure Dynamics

This section presents an analysis of how different players compute the value of a certain plant. A player computes the value of a plant differently in the two different types of market, but the main algorithm is the same. Further, Theorem 7.6 shows that, when compared to the single-clearing mechanism, the multi-clearing mechanism leads to less concentrated markets. Moreover, Theorem 7.3 shows that the forces behind portfolio of plant synergies lead to increasing market concentrations, and thus, it would seem that, although liberalised, electricity markets need regulation.

7.2.1 The Value of Electricity Plant

Assume that a player, at a certain time, holds a model of the world mapping each one of his possible actions into a possible trade then, in this way, this model summarises the trades available to this player. Therefore, this player can infer which trades are possible and the value of each plant in the new portfolio. However, even a player that has the capability of looking ahead, trying to identify trading opportunities, can fail to perceive the "portfolio value" of a plant.

Next, Definition 7.3 presents the value of a plant, showing that it is misleading to accept the common intuition that the value of a plant is a function of the cash flows it generates. A plant that generates negative cash flows may still have a positive contribution to the value of the portfolio, and thus a player may pay a positive price for it, or, if selling, he may receive a positive price for that plant.

Define the following variables for a plant *j* owned by a player *i*:
V(j,i): Total Value of plant j,
OP(j,i): Operational Profit of plant *j*,
PC(j,i): Portfolio Contribution of plant *j*,
$C_{j,L}$: total cost of plant *j* in market *L,*
$Q_{j,L}$: load of plant *j* in market *L,*
$K_{j,L}$: available capacity of plant *j* for market *L,*
$Q_{(-j,i),L}$: load of player *i*'s plants, with exception of plant *j*, in market L,
$Q_{-i,L}$: load of all the plants *not* owned by player *i*
$P_{L,F}$: market price in market *L* when a plant *j* offers its full capacity in *L*,
$P_L = A_L - \alpha_L . \sum_j Q_j$: inverse residual demand function in a market *L*.

DEFINITION 7.1: *For a certain player i and plant j:*

(a) $OP(j,i)=\sum_L D_L\left(A_L-\alpha_L.Q_{-i,L}-\alpha_L.Q_{(-j,i),L}-\alpha_L.Q_{j,L}\right).Q_{j,L}-\sum_L D_L.C_{j,L}$;

(b) The operational profit of player i is:

$$OP(i)=\sum_j\sum_L D_L\left(A_L-\alpha_L.Q_{-i,L}-\alpha_L.Q_{(-j,i),L}-\alpha_L.Q_{j,L}\right).\left(Q_{j,L}+Q_{(-j,i),L}\right)-$$
$$-\sum_j\sum_L D_L.\left(C_{j,L}+C_{(-j,i),L}\right)$$

The portfolio contribution of a plant represents the profit variation due to a reduction of the output of this plant when compared to its full capacity. Let $\max\left(\Delta Q_{j,L}\right)$ represent the maximum possible increase of generation from plant j for market L, given the capacity constraints.

DEFINITION 7.2 (Portfolio Contribution): *Defining the current market price in market L as* $P_L=A_L-\alpha_L.Q_{-i,L}-\alpha_L.Q_{(-j,i),L}-\alpha_L.Q_{j,L}$, *and the price that would result from offering the remaining capacity in that market as* $P_{L,F}=A_L-\alpha_L.Q_{-i,L}-\alpha_L.Q_{(-j,i),L}-\alpha_L.Q_{j,L}-\alpha_L.\max\left(\Delta Q_{j,L}\right)$, *it follows that*

$$PC(j,i)=\sum_L D_L.\left(P_L-P_{L,full}\right).Q_{(-j,i),L}\text{, and thus}$$

$$PC(j,i)=\sum_L \alpha_L.\max\left(\Delta Q_{j,L}\right).Q_{(-j,i),L}.D_L.$$

Thus, the definition of economic value of a plant j (Definition 7.3) can now be presented. The economic value of plant j owned by a player i, $V(j,i)$, is the sum of the operational profit with the portfolio contribution.

DEFINITION 7.3: $V(j,i) = OP(j,i) + PC(j,i)$.

Hence, the economic value of a plant is the maximum price a player is willing to pay for it. The economic value of a plant is a function of the operational profit generated by that plant plus the portfolio contribution of that plant. Joskow and Kahn (2002) present a similar concept by examining the portfolio implications of capacity withholding. The concepts developed in this chapter of the economic value of a plant and portfolio contribution extend Joskow and Kahn's analysis (which is focused on the supply function slope) by decomposing the economic value and portfolio contribution into their finer components. This allows the analyses of how plant trade relates to profit maximisation through capacity withholding and *vice versa*. Next, Figure 7.1 illustrates the concepts of economic value and profit contribution of a plant.

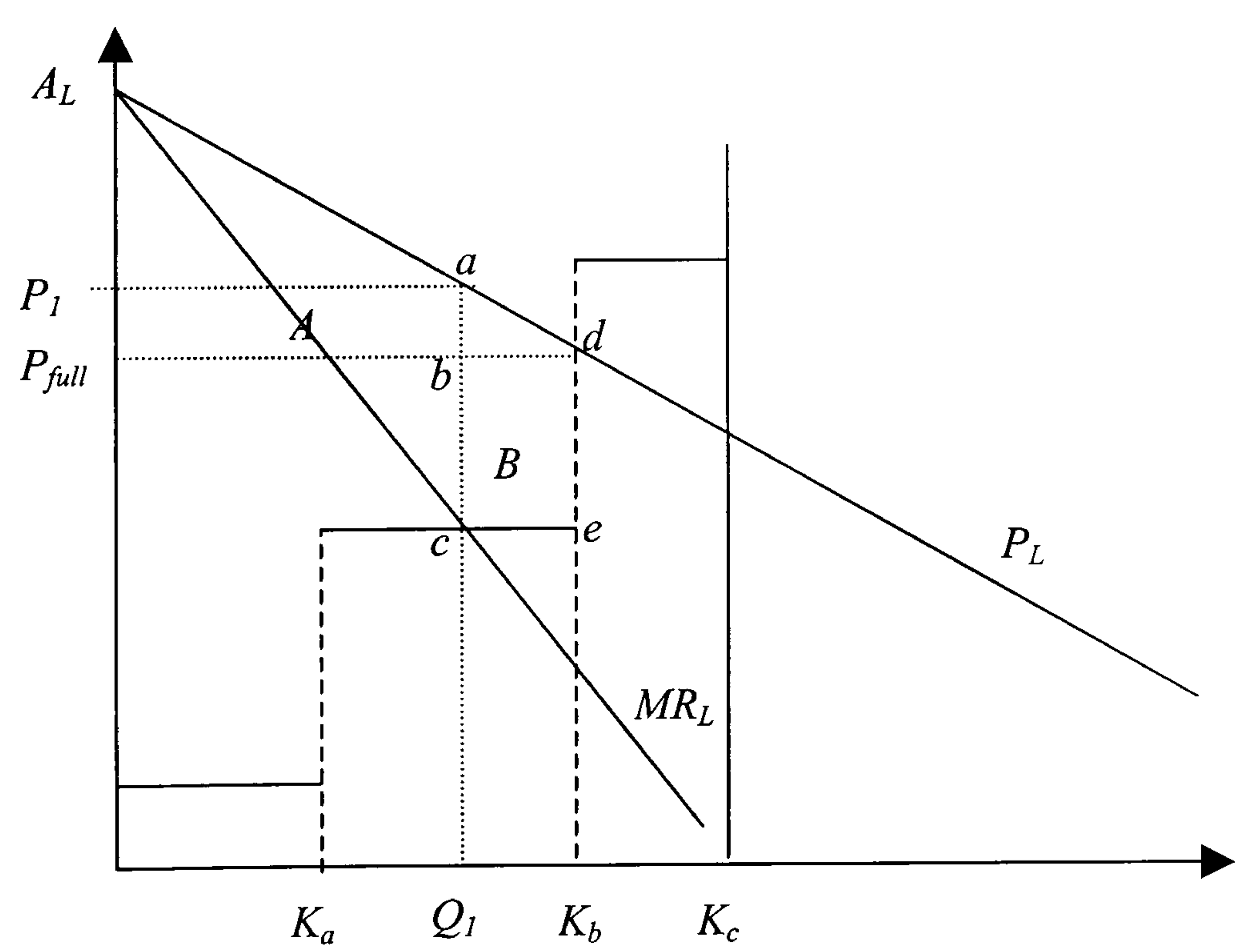


FIGURE 7.1: Example Illustrating the Concept of Portfolio Contribution.

In this example, suppose that a player i holds three plants a, b, c, which have marginal costs MC_a, MC_b and MC_c, and available capacities $K_{a,L}$, $K_{b,L}$ and $K_{c,L}$, respectively. Further, let $P_L = A_L - \alpha_L . \sum_j Q_j$ and $MR_L = A_L - 2.\alpha_L . \sum_j Q_j$ describe, respectively, the inverse residual demand function and the marginal revenue of player i, in a market L.

In the example, player i generates Q_1 units at a price P_1, and therefore he reduces the operational value of plant b by area B and increasing the value of his portfolio by A. Area A represents the portfolio contribution of plant b.

Next, Proposition 7.3.a) shows that when a player i sells the full capacity of a plant j, the operational profit of this plant is an upper bound on its value. Further, Proposition 7.3.b) shows that when a player i withholds some of the capacity of a plant j, the portfolio contribution of this plant is positive only in markets where the clearing price is higher than this plant's marginal cost (MC_j).

PROPOSITION 7.3: *(a) Let* $Q_{j,L} = K_{j,L}$, *then* $PC(j,i) = 0$. *(b) Let in a market L* $Q_{j,L} < K_{j,L}$ *then* $PC(j,i) > 0$ *only if* $MC_j < P_L$.

PROOF: (a) Let $Q_{j,L} = K_{j,L}$, it follows that $\max\left(\Delta Q_{j,L}\right) = 0$ by Definition 7.2, and therefore $PC\left(j,i\right) = \sum_L \alpha_L . \max\left(\Delta Q_{j,L}\right) . Q_{(-j,i),L} . D_L = \sum_L 0 = 0$. Since, by Definition 7.3, $V\left(j,i\right) = OP\left(j,i\right) + PC\left(j,i\right)$, it follows that $V\left(j,i\right) = OP\left(j,i\right)$, and therefore, $OP\left(j,i\right)$ is an upper bound on the value of plant j. (b) Let $Q_{j,L} < K_{j,L}$ and assume that for any market L $MC_j \geq P_L$. Thus, by Definition 7.2, $P_{L,F} \leq P_L$ and $MC_j \geq P_{L,F}$. Therefore, by the definition of marginal profit and Definition 7.1, $Q_{j,L} = 0$ and

$\max\left(\Delta Q_{j,L}\right)=0$, which, by Definition 7.2, implies that $PC_j = 0$. Hence, this proves by contrapositive that $PC\left(j,i\right)>0$ only if $MC_j < P_L$ for at least one market *L*. Q.E.D.

Next, Proposition 7.4 shows that a plant has different value in different types of portfolio. Particularly, a marginal plant is more valuable in bigger portfolios.

PROPOSITION 7.4: *In a Cournot game, the value of a plant j is a function of the portfolio of plant* $C_1 < .. < C_j < .. < C_P$ *to which it belongs. (a) In portfolios with larger total output, the portfolio contribution of a marginal plant to the portfolio is greater than in portfolios with small output. (b) Thus, in portfolios with larger total output the generation of a marginal plant tends to be lower than in portfolios with smaller output.*

PROOF: By Definition 7.3 the value of a certain plant $V\left(j,i\right)=OP\left(j,i\right)+PC\left(j,i\right)$ can be decomposed into Operational Profits and Portfolio Contribution. Assume $MC_j < P_L$. (a) By definition $PC(j,i)=\sum_L D_L.\left(P_L - P_{L,F}\right).Q_{(-j,i),L}$, and thus the larger $Q_{(-j,i),L}$ is the larger is the profit contribution of plant *j*. (b) By Definition 7.1.b) and by deriving the optimality condition for a certain output $Q_{j,L}$, it follows that the marginal loss of a portfolio is *Loss of Portfolio* = $-2\alpha_L D_L Q_{(-j,i),L}$. Thus, by the optimality conditions, the larger $Q_{-j,i}$ is the smaller is the total output of a marginal plant *j*. Q.E.D.

7.2.2 Market Structure Dynamics

From the previous sections it follows that the main market drivers are captured by Theorem 7.2, which shows that a player that buys a more efficient plant can profitably withhold capacity of his more expensive plants, and Proposition 7.2 which shows that the value of a plant is a function of the type of portfolio to which it belongs.

From Theorem 7.2 it follows that a player buys a plant that has a positive value, due to its operational profit or due to its portfolio contribution. On the other hand, a player is willing to sell a plant only to a player that values it more than he does.

Let i_s and i_b represent the seller and a buyer of a plant j, and $V(j,s)$ and $V(j,b)$ represent the respective valuations of this plant. Further, for a player a and a plant j generating $Q_{j,a}$, let $OP_{j,a}(Q_{j,a})$ and $PC_{j,a}(Q_{j,a})$ represent, respectively, the operational profit and the portfolio contribution of this plant.

Next Theorem 7.3 shows that for any plant j traded by two players s and b the weighted average of the buyer's output is higher than the weighted average of the seller's output.

THEOREM 7.3: *For any plant j,* $V(j,b) > V(j,s)$ *only if*

$$\sum_L D_L.(P_L - P_{L,F}).Q_{(-j,b),L} > \sum_L D_L.(P_L - P_{L,F}).Q_{(-j,s),L}\ .$$

PROOF: First assume that the trade occurred and therefore $V(j,b) > V(j,s)$, which is equivalent to $OP_{j,b}(Q_{j,b}) - OP_{j,s}(Q_{j,s}) + PC_{j,b}(Q_{j,b}) - PC_{j,s}(Q_{j,s}) > 0$. Since for the seller, by definition of optimum behaviour, $OP_{j,s}(Q_{j,s}) + PC_{j,s}(Q_{j,s}) \geq OP_{j,s}(Q_{j,b}) + PC_{j,s}(Q_{j,b})$, it follows that $OP_{j,b}(Q_{j,b}) - OP_{j,s}(Q_{j,b}) + PC_{j,b}(Q_{j,b}) - PC_{j,s}(Q_{j,b}) > 0$.

As $OP_{j,b}(Q_{j,b}) = OP_{j,s}(Q_{j,b})$, it follows that $PC_{j,b}(Q_{j,b}) - PC_{j,s}(Q_{j,b}) > 0$. Replacing the profit contribution by its definition, it follows that $\sum_L D_L.(P_L - P_{L,F}).Q_{(-j,b),L} > \sum_L D_L.(P_L - P_{L,F}).Q_{(-j,s),L}$. Q.E.D.

Given the expectations regarding other players' behaviour, the only possible plant transactions are the ones in which the buyer expects to reduce the output of the plant bought. This is shown in Corollary 7.1 and Theorems 7.4 and 7.5.

Next, Theorem 7.4 shows that if for any plant *j* the buyer's residual output is higher than the seller's residual output, in the markets where *j* can sell, than plant trade implies a reduction of *j*'s output.

THEOREM 7.4: *For any plant j such that* $V(j,b) > V(j,s)$: $Q_{(-j,b),L} > Q_{(-j,s),L}$ if and only if $Q_{(j,b),L} < Q_{(j,s),L}$.

PROOF: The profit of any player in a market *L*, for any player *i* owning P plants, is $\pi_{i,L} = \sum_{g=1}^{P} (P_L - C_{g,L}).Q_{(g,i),L}.D_L$. By the optimality conditions presented in the proof of

Proposition 7.2 and adapting the notation to deal with a buyer and a seller, it follows that the optimal load for plant j is $Q_{(j,i),L} = -\frac{1}{2}\sum_{g=1}^{j-1} Q_{(g,i),L} + \frac{1}{2\alpha_L} A_L - \frac{1}{2\alpha_L} C_{j,L}$, or equivalently, $Q_{(j,i),L} = -\frac{1}{2} Q_{(-j,i),L} + \frac{1}{2\alpha_L} A_L - \frac{1}{2\alpha_L} C_{j,L}$. Thus, $Q_{(j,b),L} < Q_{(j,s),L}$ is equivalent to $-\frac{1}{2} Q_{(-j,b),L} + \frac{1}{2\alpha_L} A_L - \frac{1}{2\alpha_L} C_{j,L} < -\frac{1}{2} Q_{(-j,s),L} + \frac{1}{2\alpha_L} A_L - \frac{1}{2\alpha_L} C_{j,L}$ and hence $Q_{(-j,b),L} > Q_{(-j,s),L}$. *Q.E.D.*

Further, Theorem 7.5 proves that a player whose residual output is lower than the current owner's residual output cannot buy plant *j*.

THEOREM 7.5: *For any plant j such that* $Q_{(-j,b),L} < Q_{(-j,s),L}$: $V(j,b) < V(j,s)$.

PROOF: The proof follows by contradiction. Suppose that plant j was traded between players b and s and therefore $V(j,b) > V(j,s)$. Then, as $Q_{(-j,b),L} < Q_{(-j,s),L}$ from Theorem 7.4 it follows that $Q_{(j,b),L} > Q_{(j,s),L}$. As $V(j,b) > V(j,s)$ is equivalent to $OP_{j,b}(Q_{j,b}) + PC_{j,b}(Q_{j,b}) > OP_{j,s}(Q_{j,s}) + PC_{j,s}(Q_{j,s})$, by the optimality conditions it follows that $OP_{j,b}(Q_{j,b}) + PC_{j,b}(Q_{j,b}) > OP_{j,s}(Q_{j,b}) + PC_{j,s}(Q_{j,b})$ and thus, by definition of operational profit $PC_{j,b}(Q_{j,b}) > PC_{j,s}(Q_{j,b})$. However, as $Q_{(-j,b),L} < Q_{(-j,s),L}$ and $Q_{(j,b),L} > Q_{(j,s),L}$ from definition of profit contribution it follows that $PC_{j,b}(Q_{j,b}) < PC_{j,s}(Q_{j,s})$. Hence, if $Q_{(-j,b),L} < Q_{(-j,s),L}$ then $V(j,b) < V(j,s)$, therefore b and s do not trade plant *j*. Q.E.D.

Thus, it follows from Theorems 7.4 and 7.5 that every trade implies a reduction of the output of the plant being traded: this is shown in Corollary 7.1.

COROLLARY 7.1: *For any plant j:* $V(j,b) > V(j,s)$ *only if* $Q_{(j,b),L} < Q_{(j,s),L}$.

PROOF: From Theorem 7.5 it follows that in states of the industry where $Q_{(-j,b),L} < Q_{(-j,s),L}$ there is no trade as $V(j,b) < V(j,s)$. Moreover, Theorem 7.4 specifies that if $Q_{(-j,b),L} > Q_{(-j,s),L}$ and trade does happen, i.e., $V(j,b) > V(j,s)$, then $Q_{(j,b),L} < Q_{(j,s),L}$. Q.E.D.

Next, Theorem 7.6 shows that the trading dynamics leads to markets more concentrated within the single-clearing mechanisms. While capacity withholding within the single-clearing mechanism rewards all the plants selling at a certain time, capacity withholding within in the multi-clearing mechanism, at a certain time, only benefits the plants selling in the market from which the capacity is withheld.

THEOREM 7.6: *The multi-clearing mechanism, in the long run, leads to a level of market concentration and electricity prices lower than the ones achieved by the single-clearing mechanism.*

PROOF: Let $Q_{(-j,i),t}$ represent the residual quantity sold by player *i* at time *t*, and let $Q_{(-j,i),L}$ stand for the quantity sold by player *i* in market *L*, at time *t*. By Definition

7.2, $PC(j,i)=\sum_{L}\alpha_L.\max\left(\Delta Q_{j,L}\right).Q_{(-j,i),L}.D_L$, it follows that a plant only has a positive portfolio contribution in markets where some other plants of the same player are also selling, i.e., $Q_{(-j,i),L}>0$. Since, at any given time, by definition of single clearing mechanism $Q_{(-j,i),t}=Q_{(-j,i),L}$ and by definition of multi-clearing mechanism $Q_{(-j,i),t}=\sum_{L}Q_{(-j,i),L}$, it follows that $Q_{(-j,i),L}\leq Q_{(-j,i),t}$, therefore by Proposition 7.4.b) there is less pressure for market concentration and capacity withholding. Q.E.D.

7.3 Plant Trading Game

The plant trading game is a multi-stage game of incomplete information where in the first stage each player chooses the amount of capacity he wants to hold from each different technology and, in a second-stage, he specifies the quantity of generation he wants to sell in the market. It is noteworthy that this game is not just a repetition of a single-stage game: the structure of the market changes as players buy and sell plants, thus the mapping of payoffs of the single-stage game changes as well.

Thus, the plant trading game represents a search mechanism in the space of possible market structures. In this game, the search dynamics is a function of the strategic decisions of each player in the industry.

7.3.1 Complexity of the Plant Trading Game

In a game with N players and M plants, let the vector $\Omega = \{(W_1, K_1, C_1), \ldots, (W_i, K_i, C_i), \ldots, (W_M, K_M, C_M)\}$ describe the state of the game, i.e., the ownership structure of the industry. In Ω, the triples (W_i, K_i, C_i) represent the owner (W_i), the capacity (K_i), and the cost (C_i) of a plant i.

Further, let the vector A^i represent player i's actions and $\boldsymbol{A} = \{A^1, A^2, \ldots, A^N\}$ represent an ordered vector of the actions of all the N players. Note that any given instance of $\boldsymbol{A}$ represents a transition between the states of the industry (possibly a transition with no trade, or with more than one simultaneous trade).

Next, Proposition 7.5 analyses the complexity of the plant trading game by computing the number of states of the industry and the number of transitions between states (which also include transitions in which *no* trade occurs).

PROPOSITION 7.5: *In the plant trading game, each player, at each stage of the game, can play M+1 different actions. Thus there are* $(M+1)^N$ *transitions between states and* N^M *possible states of the industry.*

PROOF: At every stage of the game: (a) A player may try to buy a plant that he does not own, sell a plant that he owns, or keep the same portfolio. Therefore, a player has one possible action per each one of the M plants in the industry, and an extra one which is to do nothing. Hence, he can play $M+1$ possible actions. (b) The number of possible transitions between states is the Cartesian product of the possible actions of each player, hence $(M+1)^N$. (c) The state of the industry is described by the ownership of each plant. Since each plant may be owned by each one of the N players, the number of possible states of the industry is the Cartesian product of the possible owners of each plant, hence N^M. Q.E.D.

Please note that the number of transitions between states and the number of possible states of the industry are an exponential function of the number of players and the number of plants, respectively. The implication of this is striking: in order for a player to analyse all possible transitions between states S stages ahead, he has to analyse $(M+1)^{SN}$ possible combinations. Moreover, as plant trading implies a bilateral agreement between a buyer and a seller for the same plant, this is a very hard coordination problem.

Therefore, given the complexity of this problem, it is difficult for a regulatory authority or company in this market to implement or compute the "optimal market structure".

7.3.2 Plant Trading Game

As presented in Table 7.1 there are four main stages in the plant trading game: Identification, Adaptation, Plant Trading and Update State of Game.

TABLE 7.1: Plant Trading Game

While the last iteration is not reached:

1. Identification,
2. Adaptation,
3. Simulate Plant Trading,
4. Update State of Game.

During Identification, each player infers a model representing how the system is behaving, and identifies the plants that will be offered most probably in the next trading round. In Adaptation, each player computes his best response to the inferred model by using an adaptive best behaviour. Then, possibly, two of the players trade a plant. Finally, the algorithm updates the state of the game.

7.3.3 The Identification Algorithm

Table 7.2 presents the Identification algorithm that works as follows. A player infers a model of how the system behaves by keeping in memory the results of each one of his actions $\left(A_t^i\right)$ in the last K periods. These results, D^i, are trade-possible (1) or not trade-possible (0). Further, a player is able to infer the results of actions that he did not take $\left(\Sigma^i \setminus A_t^i\right)$ by analysing if-then-else scenarios. The difference between the latter and the former is that actions actually submitted to an auction, A_t^i, influence the perception the other players hold on the system's behaviour, while actions not submitted do not.

Each player i then updates a Plausibility Table T_t^i, which forecasts, for every possible action, if there is a possibility of trade (1) or not (0). This plausibility is computed using a cut-off parameter θ, and acts as passive inertia[33] that discards actions that are not plausible. Thus, given the K-string of possible-events associated with each possible action, a player computes the percentage of time it would be possible for a trade to have happened, $p_{i,t}^a$, and, if $p_{i,t}^a \geq \theta$ this action is considered to be a plausible trade.

[33] Passive inertia is the behaviour of a player who decides not to act due to a lack of confidence in the model learned, instead waiting for further information. See Chapter 6.

TABLE 7.2: Identification Algorithm

D^i: Perceived outcomes of the player's actions in the path of his automaton, $D \equiv \{0,1\}$

Σ^i : Set of actions available to player i

A_t^i : Set of actions actually bid by player i, in state t, with size W: $A_t^i \subseteq \Sigma^i$

T_t^i : Plausibility Table, a one-dimensional table of dimension M *(number of plants)*

θ : Plausibility cut-off parameter

$S \equiv$ All prefixes of D^i with a length less of equal than $K > W$

At stage zero initialise (S, T_0^i): $\forall a^i \in \Sigma^i, s(a^i) = [1,1,...,1], T_0^i(s(a^i),\theta) = 1$.

1. At any given stage t:

1.a) For each possible action update the string of past perceived outcomest

$$\forall a^i \in \Sigma^i, s_t^i = \phi\left(s_{t-1}^i, D_t^i\left(a_t^i\right)\right)$$

1.b) Compute $p_{i,t}^a$ the percentage of time that each action would be successful

Let $d_j \in s_t^i$ represent a perceived outcome in string s_t^i, such that $d_j \in \{0,1\}$.

$$\forall i, \forall a^i \in \Sigma^i, p_{i,t}^a = \frac{\sum_{j=1}^{K} d_j}{K}$$

1.c) Let $\tau_{j,t}(a_t^i)$ represent the perceived outcome of action a_t^i, such that $\tau_{j,t}(a_t^i) \in \{0,1\}$.

$$\forall i, \forall \tau_{j,t}(a_t^i) \in T_t^i, \tau_{j,t} = \Phi^i\left(p_{i,t}^a, \theta\right).$$

2. The Update operator

Let $D_t^i\left(a_t^i\right)$ represent the expected outcome of action a_t^i, and let $s_{t-1}^i = [d_1, d_2, ..., d_K]$ represent the vector of the past outcomes of action a^i :

$$s_t^i = \left[d_2, ..., d_K, D_t^i\left(a_t^i\right)\right].$$

3. The Forecast operator

$$\Phi(p,\theta) = \begin{cases} 1 \leftarrow p \geq \theta \\ 0 \leftarrow p < \theta \end{cases}$$

Further, this identification process obeys two rules necessary for the rational behaviour (defined in Chapter 6) of a certain player. The first rule is consistency: the model identified by each player has to be consistent, i.e., the same action, in a certain state, always leads to the same new state. The second rule is completeness, i.e., a player builds a model that forecasts the outcome of every possible action.

7.3.4 Adaptation Algorithm

Table 7.3 presents the Adaptation algorithm that applies three principles in order to model rational behaviour: passive inertia, active inertia, and best response behaviour. Passive inertia reflects the cost of changing. Active inertia represents the conduct of a player that imposes his behaviour to others, also known as teaching behaviour[34]. Finally, best response behaviour is the attitude of a player maximising the value of his portfolio in stable environments.

TABLE 7.3: Adaptation Algorithm

Σ^i : Set of actions available to player i

A_t^i : Set of W actions actually bid by player i, in stage t, *such that* $A_t^i \subseteq \Sigma^i$

T_t^i : Plausibility Table, vector of dimension M *(number of plants)*

Ω_t : State of the industry at time t

ρ_i : Discount factor for agent i, $0 \le \rho_i \le 1$

V_t^i : Value of i's portfolio at time t

r: random generated number from a uniform distribution, such that $r \in [0,1]$

w^i: active inertia variable, such that $w^i \in [0,1]$

S: number of steps of look-ahead

[34] The concept of active inertia is presented in Chapter 6.

1. Each player i decides to adapt

 1.a) Applies Active Inertia principle, for a given w^i

$$\begin{cases} Z_t^i = BR\left(\Omega_t, T_t^i, \rho_i\right) \leftarrow r \geq w^i \\ Z_t^i = A_{t-1}^i \leftarrow r < w^i \end{cases}$$

 1.b) Algorithm Best-Response $Z_t^i = BR\left(\Omega_t, T_t^i, \rho_i\right)$:

 Compute the optimal policy, Z_t^i:

$$\forall t = 1, \dots, S,$$

$$Z_t^i = \arg\max_{a_t^i}\left[u\left(\Omega_t, a_t^i\right) + \rho_i V_{t+1}^i\left(\Omega_{t+1}, T_{t+1}^i\right)\right]$$

$$s.t.$$

$$\forall \tau_{j,t+1} \in T_{t+1}^i, \tau_{j,t+1} = \delta^i\left(\tau_{j,t}, a_t^i\right)$$

$$T_1^i = T_0^i, \Omega_1 = \Omega_0$$

$$\Omega_{t+1} = \left\{\Omega_t \setminus (a,i)\right\} \bigcup \left\{(a,j)\right\}$$

$$\delta^i\left(\tau_{j,t}, a_t^i\right) = \begin{cases} \tau_{j,t} \leftarrow j \neq a_t^i \\ 0 \leftarrow j = a_t^i \end{cases}$$

2. Complete Adaptation Model

 If $\# Z_t^i < W$

$$\text{Let } \Lambda_t^i = \left\{a_t^i : \tau_{j,t} \in T_t^i, \tau_{j,t}\left(a_t^i\right) = 0\right\}$$

$$\overline{Z_t^i} = BR\left(\Omega_t, \Lambda_t^i, \rho_i\right)$$

 else $\overline{Z_t^i} = \emptyset$

3. Define the set of actions to bid into the auction

$$A_t^i = Z_t^i \cup \overline{Z_t^i}$$

Thus, the Adaptation algorithm enables the modelling of how players learn and adapt in a dynamic environment. A player learns the automaton defining the behaviour of the environment and then he adapts his behaviour in order to maximise his long-term profit.

Given the set of plausibly successful actions $\left(T_t^i\right)$ a player i computes the set of actions $\left(Z_t^i\right)$ that would improve the value of his portfolio. However, if the number of these actions $\left(\#Z_t^i\right)$ is less than the maximum number a player can submit to the auction, he may bid some additional trading proposals $\left(\overline{Z_t^i}\right)$ that he perceives to be the most profitable, albeit having a low plausibility.

Please note that step 2 (Complete Adaptation Model) applies the active inertia concept. When a player adapts his behaviour, first he is constrained by the behaviour of the others but, at the same time, he is able to constrain their behaviour. A player that uses active inertia imposes his behaviour by two different ways. First, by choosing among the proposed trades the one he is interested in, he is able, particularly in the case where he owns the plant in discussion, to influence the plausibility of other player's actions. Second, by choosing to propose some new actions and holding to them, a player increases the plausibility of these trades, influences the other players' perceptions and "persuades" them to adapt to his own behaviour. This is the plant trading game: a battle to gain credibility, to coordinate behaviour and to gain influence, in order to do the trades that improve a player's long-term performance.

Going back to step 1.b), in order to estimate the value of state $V_{t+1}^i\left(\Omega_{t+1}, T_{t+1}^i\right)$, for each player *i*, without actually solving the Cournot game with capacity constraints (which would be solved in each possible node), each player needs a theoretical model enabling the estimation of the value of each plant in his portfolio. Fortunately, given the theory presented in sections 7.1 and 7.2, a player is able to compute the value of the planned actions.

The analysis is split into the buyer's and the seller's problem as a seller knows how to best place his plant, whereas a buyer may decide to change the market position of the plant being traded in order to adapt it to his own portfolio.

7.3.4.a) The Seller's Problem

For a given seller, the value of a plant equals the operational profit at time *t*, which he knows, plus the portfolio contribution of that plant.

Proposition 7.3 describes how the portfolio contribution is computed. A plant using in full its available capacity has null portfolio contribution and a plant that withholds capacity from a certain market has a portfolio contribution directly proportional to its owner's total load in that market.

In the single-clearing mechanism, a player knows, for each one of his plants, the quantity withheld from each market. However, in the multi-clearing mechanism, a player cannot calculate, in a straightforward way, the quantity withheld from each market. In this case, a rational player withholds capacity from the market in which that given plant receives the highest portfolio contribution.

In order to calculate the maximum possible increase of generation from plant j in market L a player identifies the potential portfolio contribution of the given plant in each market using the criterion $\textit{Loss of Portfolio} = -2\alpha_L D_L Q_{(-j,i),L}$. Thus, the quantity withheld from each market is $\max\left(\Delta Q_{j,L}\right)$. See equation (7.3).

$$\max\left(\Delta Q_{j,L}\right) = \begin{cases} K_j - \sum_L Q_{j,L} \leftarrow \forall L, L' : 2\alpha_L D_L Q_{(-j,i),L} \geq 2\alpha_{L'} D_{L'} Q_{(-j,i),L'} \\ \qquad 0 \leftarrow \textit{otherwise} \end{cases} \tag{7.3}$$

7.3.4.b) The Buyer's Problem

The evaluation of a plant by a potential buyer is slightly more complex. He needs to compute the portfolio contribution and the expected operational profit of this plant as most probably a buyer's advantage arises from the possibility of repositioning the plant.

First, a buyer computes the load of this plant in each market. This is done by maximising the operational profit of player i, (see Definition 7.1.b)), by using as decision variable the production of plant j. Again, this implies the solution of a non-linear optimisation problem every time a buyer evaluates a plant. However, the dimension of this problem is small, and its solution is easy, even if solved repeatedly.

Second, the buyer faces the same problem as the seller with regard to the computation of the portfolio contribution. Given a certain assignment of load to each market, the buyer computes the portfolio contribution of this plant using equation (7.3).

7.3.5 Simulate Plant Trading

Plant trading is organised in a single-call auction (Cason and Friedman, 1997). Table 7.4 describes the trading auction algorithm. There is a separate auction for every plant, simultaneously. The algorithm also calculates the transaction price for the plant traded.

First, a trade is possible only if simultaneously there are one or more buyers and a seller, and the price offered by the buyers is higher than the seller's bid. Second, for each plant a, the algorithm computes the transaction price at time t ($P_{a,t}$). By calculating the simple average of the seller's bid price and the buyers' highest offered price. Then, $P_{a,t}$ equals the maximum of this simple average and the second highest offered price. Third, admissible trades are only those ones where the transaction price is positive, as the seller would never pay to sell (due to the implicit option of closing-down a plant).

TABLE 7.4: The Trading Auction

a: Asset being auctioned
i, j: players offering (attempting to sell) or bidding (attempting to buy) assets in an auction
$P_{a,t}$: Transaction price of asset a at time t
$B_{a,i}$: Price bid by player i attempting to buy asset a
$O_{a,i}$: Price offered by player i attempting to sell asset a
T_a: Set of all possible trades for asset a
B_a: Set of all acceptable bids for asset a
T: Set of the all winning trades (at the most one per asset)

1. For every asset a find T_a

$$T_a = \{(B_{a,i}, O_{a,j}) : i \neq j, B_{a,i} > O_{a,j}\}$$

$$B_a = \{B_{a,i} : (B_{a,i}, O_{a,j}) \in T_a\}$$

2. Find T:

For every asset a find the winning trade $(B^+_{a,i}, O^+_{a,j})$

If $T_a \neq \emptyset$ then $O^+_{a,j} = O_{a,j}$ and $B^+_{a,i} = \sup B_a$

$$T = \bigcup_a (B^+_{a,i}, O^+_{a,j})$$

3. Find the asset to be traded $(B^*_{a,i}, O^*_{a,j})$

Let g stand for a function from T into $\mathbb{R}$:

$$g = \{((B_{a,i}, O_{a,j}), G_a) \in T \times \mathbb{R} \mid G_a = B_{a,i} - O_{a,j}\}$$

The asset to be traded is the one with the largest difference between offer and bid prices

$$(B^*_{a,i}, O^*_{a,j}) = \underset{(B_{a,i}, O_{a,j})}{argmax}\ g$$

4. Compute the transaction price

Let $B^{++}_{a,z}$ represent the second highest bid for asset *a:*

$$B^{++}_{a,z} = \sup\left\{B_a \setminus B^*_{a,i}\right\}$$

$$P_{a,t} = \max\left(\frac{B^*_{a,i} + O^*_{a,j}}{2}, B^{++}_{a,z}\right)$$

Next, after computing the transaction price of each plant, the auctioneer chooses which transaction to take place at time t. It is possible to have more than one trade per iteration, and indeed, this would not hurt any of the theoretical properties of the model. However, this would imply that the jumps between successive states of the industry would be wider and the evaluation error for each plant would be higher. These errors are due to the uncertainty created by simultaneous moves of the players in the game. Moreover, the bigger the plants traded the larger these errors would be. Thus, the assertion that only one plant is traded at a time only implies a smoother adjustment trajectory.

7.3.6 Update State of the Game

After any successful trade, the algorithm computes a new state of the game as the portfolio of two players and, most likely, the equilibrium prices and the output of the Cournot game would change. Moreover, even if there is no trade, the probabilities associated with active inertia principle still need to be updated. Table 7.5 describes the algorithm updating the state of the game.

TABLE 7.5: Update State of the Game

---

a: Any given asset that may be auctioned
i, j: players offering (selling) or bidding (attempting to buy) assets in an auction
(a,i): Asset a is owned by player i
$P_{L,t}$: Electricity price in market L, at time t
Ω_t : State of the industry at time t
$mg(a)$: marginal cost of asset a
$cap(a)$: available capacity of asset a
$K(a,L)$: capacity of asset a offered in market L in the previous iteration
$mg(i,L)$: marginal cost of player i in market L
$cap(i,L)$: capacity of player i assigned to market L
w^i: active inertia variable, such that $w^i \in [0,1]$
$\sigma \in]0,1[$ is a parameter

1. Update state of the industry Ω_t

$$\Omega_{t+1} = \{\Omega_t \setminus (a,i)\} \bigcup \{(a,j)\}$$

$$w^z := \begin{cases} w^z.\sigma \leftarrow otherwise \\ 1 \leftarrow z = i,j \end{cases}$$

2. Update cost structure and capacities bid in each auction:

$\forall L, \forall i :$

$cap(i,L) := 0$

$\forall (a,i)$

$K(a,L) := 0$

2.1: For all available asset a

$$if\ mg(a) \le mg(i,L) \vee [K(a,L) = 0, mg(a) \le mg(i,L+1)]$$

$$cap(i,L) := cap(i,L) + cap(a)$$

$$mg(i,L) := max[mg(i,L), mg(a)]$$

$$K(a,L) := cap(a)$$

2.2: For the Multi-Clearing mechanism

$$cap(a) := 0$$

3. Solve Cournot game

4. Compute value of plant

$$\forall i, \forall (a,i):$$
$$OP(a,i) = \sum_{L} \left[\left(P_{L,t} - mg(a) \right) . Q(a,L) . D_L \right]$$
$$OP(i) = \sum_{(a,i)} OP(a,i)$$

The algorithm starts by replacing the name of the owner of the plant traded (player i sells plant a to player j). Then, in order to solve the Cournot game (step 3) the algorithm computes the marginal costs and capacities of each player in each auction. Finally, the algorithm computes the operational profit of each plant and player.

Each player updates the capacities and marginal costs iteratively, taking into account the past performance of each plant. A player offers in a given market the generation of every plant with a marginal cost lower than this player's marginal plant in this market[35]. Moreover, a player may offer the generation of a plant in a certain market even when its marginal cost is higher than the player's marginal cost. However, this is only possible if in the previous iteration the player did not offer the generation of this plant in this market and additionally, if the marginal cost of this plant is lower than the player's marginal cost in the following markets (assuming that the markets are organised in increasing order by the clearing price). It is noteworthy that this second condition is correct only if the player did not offer any capacity of the given plant in this market, i.e., if the generation capacity was offered but the plant did not run, then it is not offered in this market as it does not respect conditions 2.1.

A more detailed model of the supply function would include the submission of the actual step function by each player in each Cournot game. This would increase the complexity of the game as the marginal cost is non-linear: this is a mixed integer complementarity problem with equilibrium constraints (Ramos et al., 1998).

7.4 Exemplification of the Plant Trading Game

This section analyses the impact of market design on the strategic management of plant portfolios through simulating the structural evolution of the E&W electricity market under different initial conditions for the market structure. Further, this study also aims to clarify how the main attractors in this game influence a player's behaviour. The analysis proceeds by comparing the results under single and multi clearing mechanisms and showing that the theoretical conclusions presented in the previous sections hold in a computer simulation of a real market structure. The simulation refers to the structure of the E&W electricity market in 2000 (see Table 4.9), just before the introduction of the New Electricity Trading Arrangements.

These experiments simulated trading at a genset level (137 gensets) distributed among 24 different players. This leads to the existence of approximately 1.22E+189 possible states of the industry and 2.27E+51 possible transitions, at every stage of the game.

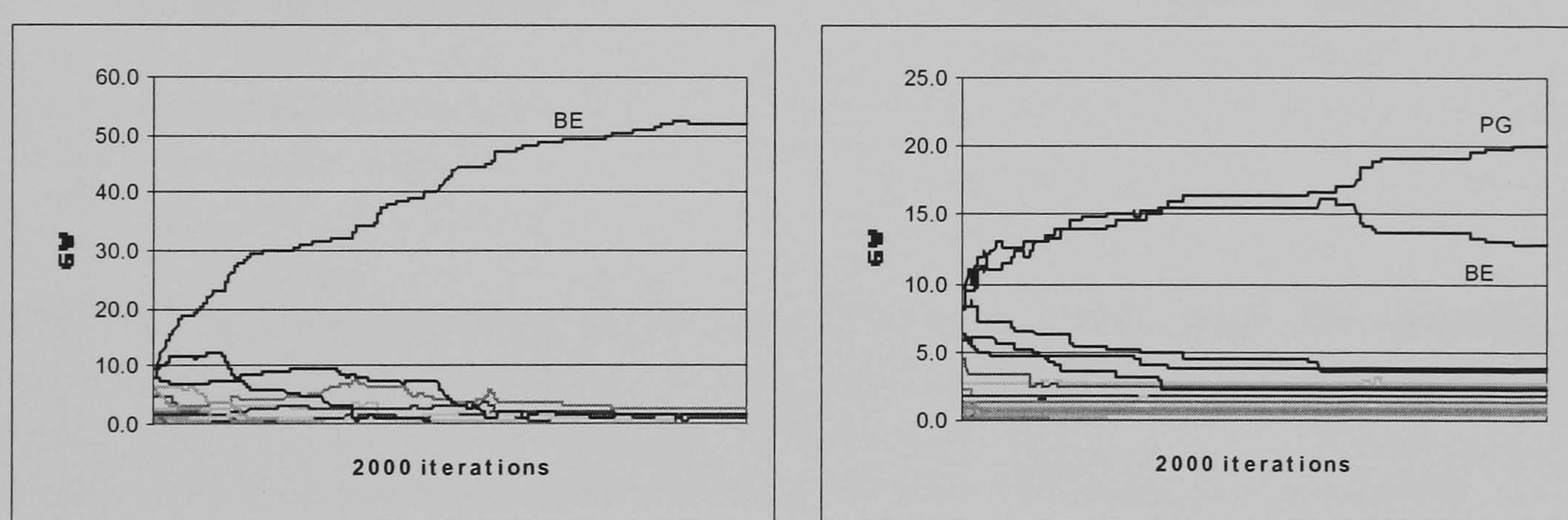
Furthermore, in these experiments the demand functions were parameterised by defining the same elasticity, prices, and traded quantities in each one of the two clearing mechanisms, in the three simulated markets (baseload, shoulder and peak). The elasticities used were 0.5, 0.35 and 0.25 respectively for the baseload, shoulder and peak market. The choice of these elasticities was fairly *ad hoc*, however it follows the elasticities used previously in the literature: Wei and Smeers (1999) use 0.4 and 0.53 for residential and industrial clients respectively, in simulating the Belgium, France, Germany and Italy market; Ramos et al. (1998) use an elasticity of 0.6 in simulating the Spanish market.

---

[35] Please note that the marginal cost of a given player is the highest one among all the plants he submits to a given market (Ramos at al., 1998; Borenstein et al. 1999, 2002).

The durations for the shoulder and peak markets were specified as 5500 and 500 hours, independently of the clearing mechanism. The duration of the baseload was specified as 8760 and 3260, respectively for the multi and single-clearing mechanisms. Further, all the experiments presented in this section simulate 2000 iterations under each different scenario.

Again, it is noteworthy that these experiments represent a very rough approximation of the real behaviour of the E&W market, aiming to illustrate the main features of the plant trading game. A more realistic simulation for policy analysis would have to look at several trading days and include regulatory constraints such as market shares, mergers control and price caps. While for theory development the economist prefers a simple model that shows the general theorems characterising the behaviour of the system, for policy analysis and design issues, the economist is faced with an engineering problem, and a detailed model would be preferred (Roth, 2002).

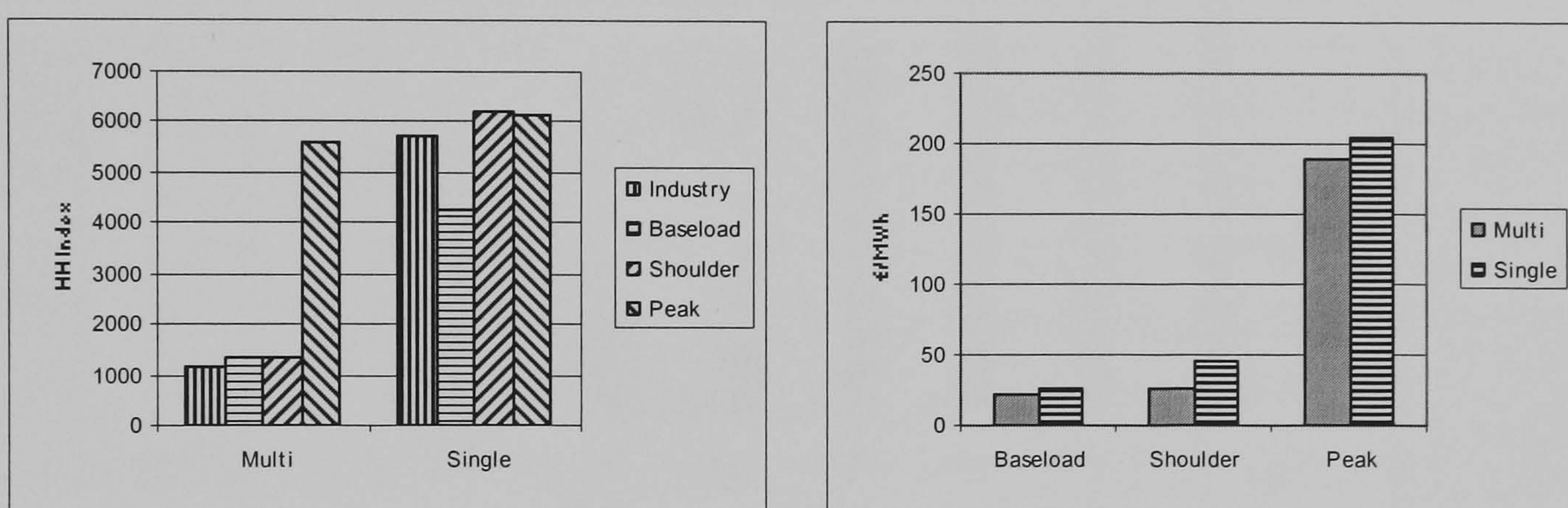


(a) (b)

FIGURE 7.2: Capacity by player. Experiment with 24 players. (a) Single-clearing mechanism. (b) Multi-clearing mechanism.

Figure 7.2 presents the results of the first set of experiments under the single and multi clearing mechanisms. As shown in Figure 7.2.a), BE becomes the incumbent player in a monopoly with a competitive fringe. This was expected, and it happened as BE is the dominant baseload generator. They bought the shoulder and peak plants in order to increase the value of their baseload portfolio, and thus became the dominant company. As illustrated in Figure 7.2.b), PG and BE are the dominant players in the multi-clearing mechanism. As expected, the experiments converged to lower concentration levels in the multi-clearing mechanism as in this case capacity withholding is less profitable than in the single-clearing mechanism.

Figure 7.3.a) shows that those under the single-clearing mechanism the concentration indices are higher than under multi-clearing mechanism. This high concentration translates itself into higher electricity prices (Figure 7.3.b)).



(a) (b)

FIGURE 7.3: Concentration and Prices. Experiment with 24 players. (a) HHI Concentration Indices: Multi (Single) represents the concentration index in the multi-clearing (single-clearing) mechanism. (b) Electricity Prices (in the Baseload, Shoulder and Peak markets) presented as a function of Clearing-Mechanism.

Since the starting values of the concentration indices under the different types of clearing mechanism were the same, these results imply that the dynamics of the single-clearing mechanism leads to higher market concentration.

This can be analysed in more detail by looking at the evolution of BE's and PG's market shares in the multi-clearing mechanism. Figure 7.4 shows that PG became a dominant player in the peak technologies (small-coal, pumped-storage, OCGT and oil) and BE became a dominant player by enlarging their dominant position in the baseload technologies (nuclear and big-coal).

The comparison of the two experiments here reported supports the theoretical claim that the multi-clearing mechanism tends to lead to lower market concentration than the single-clearing mechanism. Therefore, this seems to suggest that a single-clearing mechanism may need closer monitoring and regulatory intervention than the multi-clearing mechanism.

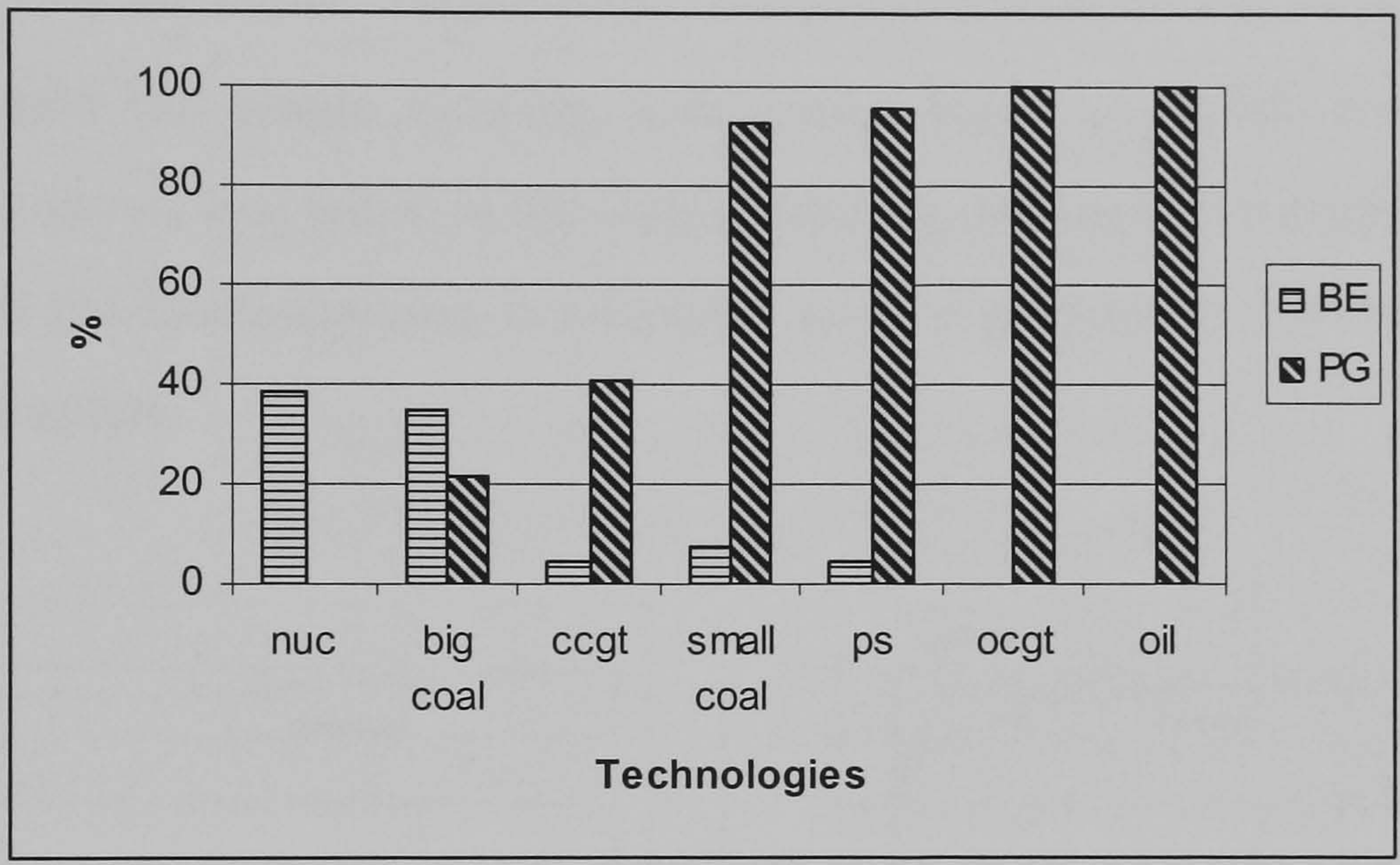


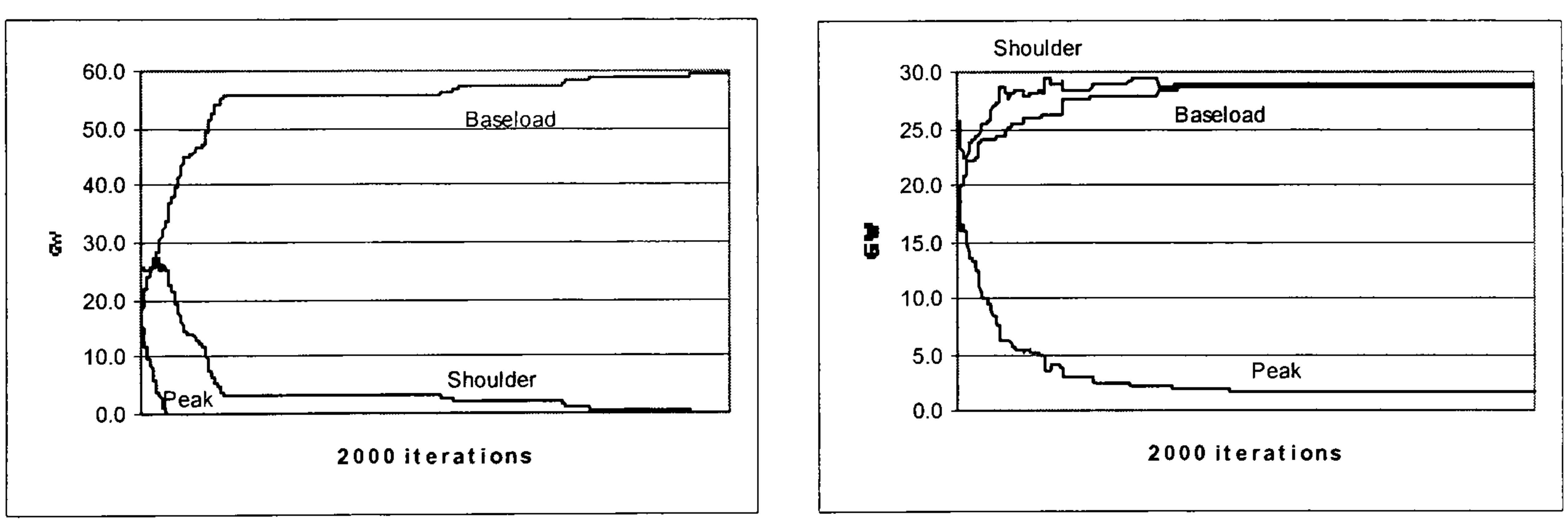
FIGURE 7.4: BE's and PG's Capacity Shares, it presents an analysis by Technology for the Multi-Clearing Mechanism.

At a technological level, the multi-clearing mechanism penalises more the baseload technologies and creates incentives for using peak-plant which would otherwise be withheld from the market in order to generate price spikes that increase the profitability of the baseload sector. These experiments also confirm that in the electricity industry, capacity withholding is a profitable activity.

Furthermore, several other experiments validate this model by testing its results under different market structures. This validation process showed that when players behave as price takers, no trading occurs as the players generate the market clearing quantities. Further, in a simulation of the monopoly situation with potential new entrants, there was no entry, and the prices were always the monopoly ones.

Next, two other sets of experiments with three players each analyse the impact of market design and market structure on equilibria.

In the two experiments presented in Figures 7.5 and 7.6, all the conditions are the same as in the first set of experiments (Figures 7.2, 7.3 and 7.4) except that the baseload, the peak and the shoulder plants were distributed among three different players only (they are called, respectively, Baseload, Shoulder and Peak). As Figure 7.5 shows, again the single clearing mechanism leads to higher concentrations. Moreover, this shows that while in the single-clearing mechanism a player becomes a monopolist, in the multi-clearing mechanism several players do survive in a more competitive structure.

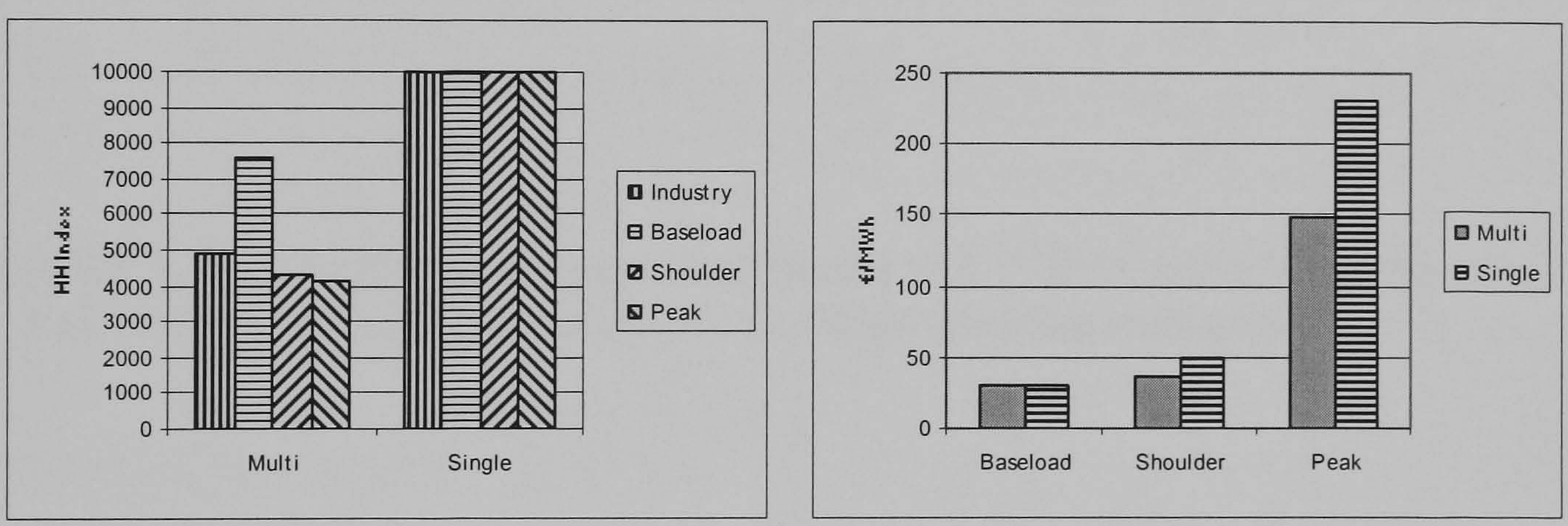


(a) (b)

FIGURE 7.5: Capacity by player. Experiment with 3 heterogeneous players (Baseload, Shoulder and Peak). (a) Single-clearing mechanism. (b) Multi-clearing mechanism.

By further looking at Figure 7.6, again the concentration indices in the single-clearing mechanism tend to be higher than in the multi-clearing mechanism.

However, the comparison of Figure 7.6 and Figure 7.3 shows that, in the multi-clearing mechanism, the dynamics behind concentration is entirely different. While in the first set of experiments the players tend to trade peak plants and therefore, the concentration index in this technology is the highest. In the second set of experiments (Figure 7.6) the concentration index of the baseload technology instead is the highest. This implies that the dynamics of trade in this experiment was completely different, as the initial values for these concentration indices were, respectively 5200, 3360 and 5000 (in Figure 7.6.a). Thus, within the multi-clearing mechanism in this experiment, the baseload player buys baseload plant to increase his profit, while in the first experiment the most traded plants belong to the peak technology. Further, note that even with only three players the multi-clearing game does not converge to a monopoly.

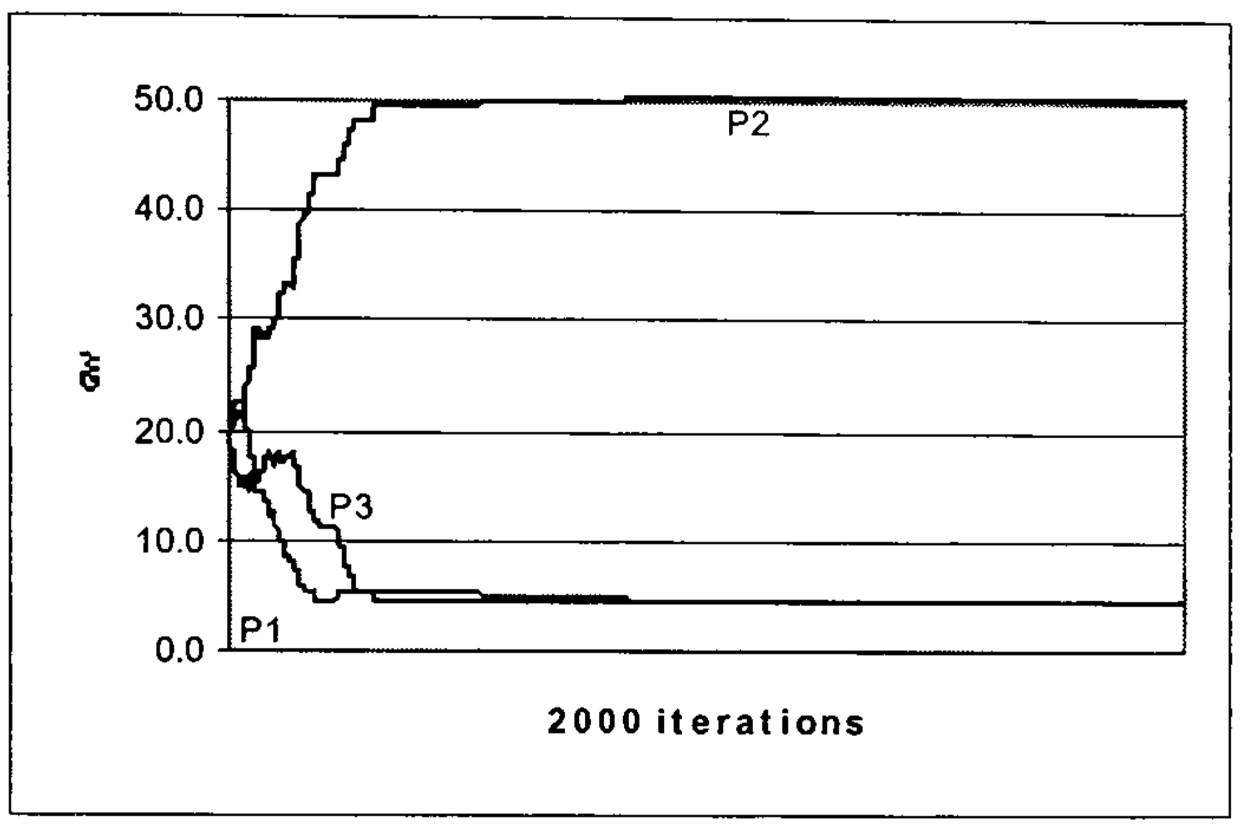


(a) (b)

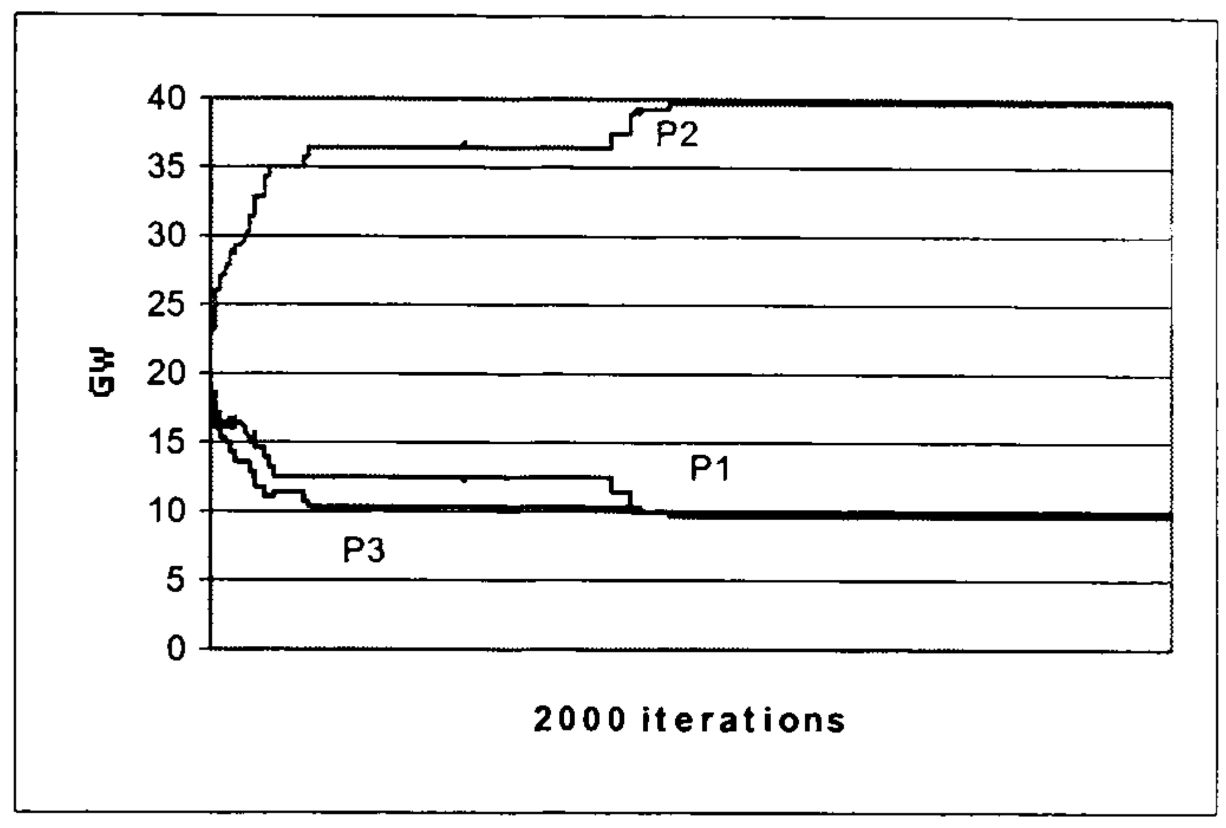
FIGURE 7.6: Concentration and Prices. Experiment with 3 heterogeneous players (Baseload, Shoulder and Peak). (a) HHI Concentration Indices: Multi (Single) represents the concentration index in the multi-clearing (single-clearing) mechanism. (b) Electricity Prices (in the Baseload, Shoulder and Peak markets) presented as a function of Clearing-Mechanism.

As a further test to the interaction of market design and the strategic management of plant portfolios, the first set of experiments was repeated once again. However, in this third set of experiments (Figures 7.7 and 7.8) the plants were assigned (one by one) to each one of three players (called P1, P2 and P3) by increasing order of marginal cost. Thus, in these experiments the initial portfolios are equal for all the three players.

Figure 7.7 represents these results, showing that in the single-clearing mechanism the structure converged to a more concentrated one than in the multi-clearing mechanism.



(a)



(b)

FIGURE 7.7: Capacity by player. Experiment with 3 homogeneous players (P1, P2, P3). (a) Single-clearing mechanism. (b) Multi-clearing mechanism.

Further, for the multi-clearing mechanism, the comparison of Figures 7.8 and Figures 7.6 shows that the market structure dynamics is different under different types of privatisation. In the experiment where the players are of different types (Figure 7.6), the state of the industry converges to a more concentrated structure. However, this is not synonymous of higher prices, as some players may have localised market power. Thus, this is the reason why the peak prices in Figure 7.6 (for the multi-clearing mechanism) are lower than in Figure 7.8. Hence, as referred to in Chapter 4, the technological flexibility introduces new opportunities to exercise market power. Again, this comparison also suggests that the dynamics of adjustment is different, as when all players are alike they trade more peak plant, while when they are of different types they trade more baseload.

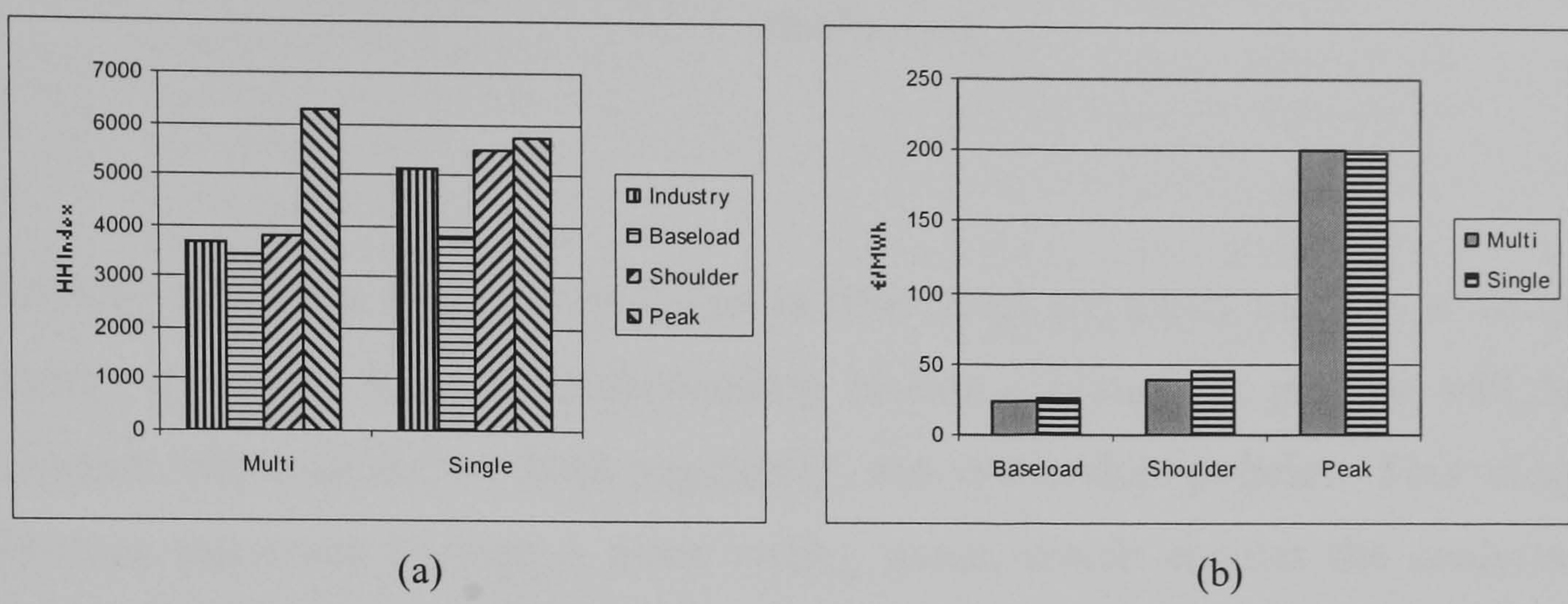


FIGURE 7.8: Concentration and Prices. Experiment with 3 homogeneous players (P1, P2, P3). (a) HHI Concentration Indices: Multi (Single) represents the concentration index in the multi-clearing (single-clearing) mechanism. (b) Electricity Prices (in the Baseload, Shoulder and Peak markets) presented as a function of Clearing-Mechanism.

Furthermore, the comparison of Figure 7.7 (7.8) and Figure 7.5 (7.6) suggests that independently of the type of the clearing mechanism, a privatisation where all the players are similar tend to lead to less concentration. However, this second scenario is very artificial as it ignores the economies of scale that still exist in some technologies such as Nuclear.

Finally, the number of trades in each game was very small when compared with the 137 gensets in the industry. Thus in the experiments with 24 players there were 168 and 105 trades respectively for the single and multi-clearing mechanism. In the experiments with three different players, there were 135 and 76 trades respectively for the single and multi-clearing mechanism. In the experiments with three equal players, there were 112 and 84 trades respectively for the single and multi-clearing mechanism.

7.5 Conclusion

The issue of market structure evolution in liberalised electricity markets is an open research question. A better understanding of this evolutionary process will have important implications on both regulatory and ownership policies. This chapter addresses this issue through a plant trading game, which enables the analysis of market structure evolution as an endogenous variable. This model provides a framework that enables an endogenous search for the possible market structure equilibria and, at the same time, gives better insights to the trajectory toward equilibrium.

Further, the search space of this problem is an exponential function of the number of plants in an industry, which means that the optimal structure of the industry and the equilibrium of this game are extremely hard to compute. Moreover, the possible number of different transitions between states is an exponential function of the number of players, meaning that the plant trading game represents a hard coordination problem where each player needs to search for bilateral trading opportunities.

Hence, in developing the plant trading game a number of behavioural properties of the players were identified, some of which challenge conventional wisdom regarding generation capacity management:

1. Capacity withholding is a profitable strategy for "portfolio players". Further, it is more profitable in the single-clearing mechanism than in the multi-clearing mechanism.

2. A player cannot profitably transfer load from a cheap to an expensive plant by withholding capacity. However, he can profit from withholding capacity of an expensive plant.

3. A player can modify the value of some of his plants by acquiring a new plant.

4. The cash flow generated by a given plant represents only a lower bound on its contribution for the value of the company. However, the cash flow of a plant generating its full capacity is the upper bound on the value of this plant.

5. Marginal plants are more valuable in portfolios with a large total output.

6. Plant trade increases market concentration.

7. The prices and concentration levels of the single-clearing mechanism tend to be higher than the ones of the multi-clearing mechanism.

8. The multi-clearing mechanism penalises more the baseload technologies and creates incentives for using peak-plant.

Therefore, at a policy level, the analysis of the plant trading game shows that market design influences not only current prices but also the long-term attractors of the industry.

Finally, the use of the Finite Automata Dynamic Game allows the simulation of bounded rationality and learning in a setting where players rarely interact.

CHAPTER 8

CONCLUSIONS

The contributions of this research are threefold:

(1) At a policy level, the contributions are the analysis of market design and market power through agent-based computational learning.

(2) At a theoretical level, the contributions are the development of a theoretical model for electricity plant valuation, and a Bertrand justification for capacity withholding.

(3) At a methodological level, the contributions are the development and analysis of a model of automata dynamics and of a model for plant trading in electricity markets. The automata game enables the modelling of evolutionary systems, even when agents seldom interact, by incorporating adaptation and learning into a game where players are boundedly rational.

The plant trading game enables the analysis of plant trading dynamics and its impact on market structure evolution.

8.1 Contributions to Policy Analysis

Chapters 4 and 5 show that agent-based computational learning can provide insights into pricing and strategic behaviour in electricity markets (even through a very simple learning algorithm). This represents the first detailed study of the New Electricity Trading Arrangements (NETA) for the England and Wales (E&W) electricity market. Further, this study gives insights into the implications of the NETA that were relevant for policy makers, before its actual introduction. The simulations in Chapter 4 allow important insights into the workings of the future market structure. First, by analysing the relation between bilateral markets and the balancing mechanism. Second, by studying the impact of forecasting errors on the efficiency of the industry. Third, by evaluating the impact of the new market rules on the performance of generators and suppliers.

Further, one last policy contribution is the analysis of the possibility of individual players (e.g. AES and BE) having the capability to profitably manipulate market prices. The simulations in Chapter 5 show that the structure of the E&W electricity market was such that prices would tend to rise significantly above marginal costs. However, these experiments support the claim that none of the players above can benefit from price manipulations, if acting alone, but when acting together they are able to profitably manipulate market prices.

Overall, an agent-based simulation of NETA enabled insight into the behaviour of the new wholesale electricity market, which was obtained in 2000 just before its introduction in 2001.

8.2 Contributions to the Theory of Electricity Markets

A first theoretical contribution is in the analysis of market power abuse and regulatory policy. Chapter 5 addresses individual market power abuse in the context of a study on the possibility of AES and BE having the capability to profitably manipulate market prices. Thus, Chapter 5 develops a stylised model of the interactions in an electricity market (modelled as a Bertrand game with capacity constraints) that shows that unilateral capacity withholding may lead to increased profits and misevaluation of market power abuse.

A second theoretical contribution is in the analysis of automata behaviour, and more specifically in the study of the conditions for rational behaviour: among other findings, showing that best response is a necessary but not a sufficient condition for rational behaviour.

A third theoretical contribution regards the issue of electricity plant trading in liberalised electricity markets. Chapter 7 develops a theoretical analysis of plant trading within a Cournot game. This model allows the identification of the players' behavioural properties, showing that capacity withholding is a profitable strategy for portfolio players and that plant trading leads to increased market concentrations. Further, this analysis proves that markets where plants are paid differently, depending on their flexibility, are less prone to market concentration, and are in general more competitive.

8.3 Methodological Contributions

Chapters 4 and 5 present a computational model for electricity trading (the NETA model) that captures the learning behaviour in a repetitive pricing game by using a simple reinforcement model. This model represents an improvement on the n-armed bandit algorithm, allowing the modelling of non-stationary environments. However, this NETA model adopts several simplifications (such as imposing rules of behaviour) in order for it to be successful.

Furthermore, there are other limitations on the use of reinforcement learning for electricity markets modelling. Due to the nature of reinforcement learning, a player needs to experiment with each one of the possible alternatives before he can learn the best policy (see Chapter 3 on the exploration *vs.* exploitation trade-off). However, this implies the taking of actions that are potentially costly, and more importantly, that would never be taken by a generation company. In this case, the NETA learning model addresses this problem by adopting rules that complement the learning algorithm, enabling the players to exhibit rational behaviour.

Therefore, before modelling the plant trading game, two problems needed solution. First, traditional evolutionary models are unsuccessful in situations where players seldom interact, as learning requires the repetition of a given action several times. Second, reinforcement learning requires players to try actions that companies do not try. The Finite Automata Dynamic Game (FADG) developed in Chapter 6, gives a solution for these two problems, modelling players as learning automata with the capability of forecasting the value (outcome) of a given action without actually taking it.

8.3.1 The Finite Automata Dynamic Game

There are two types of approaches to the automata game problem. First, a player computes the best response against his opponent assuming an exogenous set of automata from which he can choose his strategy. Alternatively, a player computes the best response against his opponent, assuming that he follows a given automaton. However, this first approach to automata games never explains the existence of the exogenous set of automata. Further, it does not explain why the opponent holds a certain automaton. Second, a player infers the automaton used by his opponent while playing against it. This second approach still assumes that the opponent has a stationary automaton, without explaining how the opponent computed this automaton.

The FADG, presented in Chapter 6, represents a new framework to model learning and adaptation in *N*-player extensive form games of incomplete information. It models the possible automata as endogenous variables, capturing the process by which a player infers a model of the other's behaviour. Hence, the FADG enables the modelling of co-evolutionary automata that learn and adapt together, reshaping the game structure.

8.3.2 The Plant Trading Game

Chapter 7 presents an analysis of the plant trading issue showing that the optimal structure of the industry and its dynamic equilibrium are extremely hard to compute. Moreover, this study shows that the plant trading game represents a hard coordination problem, due to the high number of possible combinations of trades.

Therefore, in order to model the plant trading issue, each player uses the theoretical properties of the Cournot model for plant valuation, which together with auction theory and the FADG represent the theoretical foundations of the plant trading game. Given these foundations, each player infers a model of the environment and adapts to it in a way that simulates boundedly rational behaviour in a large coordination game, simulating how players may learn in plant trading.

Chapter 7 shows, through these simulations, that single-clearing mechanisms have a dynamics that leads to higher concentration and market prices (when compared to multi-clearing mechanisms), and that baseload players tend to be more dominant in a single-clearing mechanism than in a multi-clearing mechanism. Finally, Chapter 7 proves that the coordination problem although hard, can be “solved” through simulation.

Finally, the use of the FADG allowed the simulation of bounded rationality in an environment where information is scarce, where actions are possibly irreversible, and where learning and adaptation are central to the understanding of the dynamics of the system, e.g., the plant trading game.

8.4 Concluding Remarks

In developing a simulation platform for design of wholesale electricity markets (Chapters 4 and 5), the first major option faced by a modeller is the trade-off between detailed (and dimensionality) modelling and behavioural analysis. The agent-based approach is able to develop a good balance between these two options, being able to combine detailed models of an industry structure (possibly a high number of players) with a good model of the learning behaviour in repeated games.

Additionally, a modeller faces a second trade-off between detail (accuracy and dimensionality) and analytical results. Moreover, this is just another representation of the trade-off between generalisability and the need for detail. In this case, the agent-based computational approach does not help much. The modeller himself needs to choose between developing a very detailed model of reality and capturing the basic behaviour of the phenomenon analysed. This trade-off is found repeatedly in this thesis and it is addressed differently in several problems.

Thus, the model presented in Chapters 4 and 5 gives a higher preference to dimensionality and accuracy: modelling the interactions between different markets, capturing the diversity of electricity-generation technologies, modelling an active supply side and the learning processes involved in the repeated game. However, all these complexities, if important to capture the implications of market design, implied a statistical analysis of the simulation results. This is a normal procedure for analysing models characterised by complexities that render their analytical study very hard, if not impossible.

However, in the end, it is difficult to argue that the simulation results in Chapters 4 and 5 are generalisable. In other words, they do not constitute theoretical knowledge applicable to other markets for electricity. Therefore, even if these simulation results:

1. are true for the specific market modelled,
2. represent an invaluable insight into policy making in the E&W electricity market, in 2000, specifically due to the introduction of NETA,
3. represent innovative insight, some of which is indeed generalisable,

nevertheless, due to the lack of an analytical foundation for the NETA model, this modelling technique by itself cannot prove that these claims are true.

Therefore, Chapters 6 and 7 develop the theoretical framework on which the simulation models presented in these chapters are constructed. Hence, the application of agent-based simulation to analysing the properties of an evolutionary automata game, the FADG, in Chapter 6, can be indeed generalised to any type of automata games, under the conditions presented. Moreover, any game developed within this framework shares the automata game properties. Thus, for example, the plant trading game (Chapter 7) exhibits the FADG dynamic properties. Furthermore, given that this is a stylised model capturing the problem of behavioural coordination for plant trading in electricity markets, the results presented in the plant trading game, although simulated within the E&W market structure, are generalisable to other electricity markets, under the assumptions specified.

The second part of the major trade-off faced by a researcher is to define the level of detail with which to model behaviour. If perfect-rationality seems to be an extreme case that does not capture how people and organisations behave, the modelling of bounded rationality is an issue that is very hard to address properly. Thus, Chapter 3 shows the limitations of reinforcement learning to model human behaviour, due to convergence problems, scarce information, and more importantly due to the non-stationary nature of models of learning.

The limitations of these models of learning (in which Bayesian learning is included) represented one of the major challenges faced in this thesis.

Therefore, Chapter 4 develops an improvement on the basic n-armed bandit learning algorithm, in order to prevent fast convergence, enabling modelling of a non-stationary environment. However, the NETA learning model still assumes that an agent only learns by reinforcement, he does not have any forecasting capabilities, and further assumes that this agent has enough learning opportunities to infer the value of an action by repeating it several times, and that it is not costly to try actions that are not perceived as optimal. For this reason, the FADG aims to overcome some of reinforcement learning limitations by developing a model of learning and adaptation in games where players have information constraints, where actions are possibly expensive and where players have the capability of forecasting the output of a given action without trying it. Further, note that the FADG, contrary to reinforcement learning, separates the processes of adaptation and learning into complementary algorithms.

Hence, the FADG aims to model quasi-perfect-rationality, which implies an upper bound on a player's rationality. However, it still requires several assumptions on a player's rationality that are far from being uncontroversial. Specifically, it assumes that a player can adapt optimally to his perceptions, which implies strong reasoning capabilities. Of course, these capabilities can be relaxed if a modeller finds this relaxation adequate to model the issue studied. Moreover, at the learning level, the FADG assumes that a player can infer the outcome of actions that he did not try. This may not be the case. It depends on how easily a player can acquire new information. However, a researcher can adjust a player's forecasting capabilities as required by the specific problem modelled.

Hence, the modelling of human and organisational behaviour represents a challenging task. The models of learning in games are just not good enough to capture all the complexity of human and organisational behaviour. People and organisations learn by reinforcement but they also learn by being told, they learn by induction, by deduction, by qualitative analysis, by immersion (by acting within a given social context), etc. Which type of model should a researcher choose to capture the nature of the issue he is studying? As in all scientific method, it seems that the simpler the model the better. However, in this case, it is not clear which model is the simplest or the best. This is an area where one can expect further and interesting contributions in decision science that may indeed have strong implications for game-theoretical modelling in general, and for electricity markets modelling in particular.

Finally, the liberalisation of electricity markets poses new challenges that stretch the modelling techniques to their limit. The most analysed of these issues is market design and its relation to market power. However, the work on this topic tends to represent a short-term view of this problem (of which the analysis in Chapter 4 is also guilty) failing to address the evolutionary implications that, in reality justify the privatisation process.

Thus, the dynamics of adjustment within these liberalised markets can be more important than short-term price spikes or market power abuse from generators. Therefore, Chapter 7 analyses the issue of plant trading aiming to develop theoretical foundations for modelling the dynamics of plant trading in electricity markets. However, this is just one of the first foundation stones on this topic. One should expect very important developments in the area of market structure evolution (including capacity investment) in decentralised electricity markets. More specifically, agent-based evolutionary models can bring new insight into the long-term effects of short-term policy making, possibly allowing a better evaluation of the impact of the privatisation process on consumer welfare. Further, companies in these markets can apply these evolutionary models to evaluate the long-term implications of strategic decision-making. In this regard, the plant trading game, by additionally considering the effects of regulatory intervention, new entry and technological innovation (among others) can be a very interesting tool for studying the impact of regulatory activity and strategic decisions on the evolution of market structure.

BIBLIOGRAPHY

Abbink, K., Brandts, J., McDaniel, T., 2003. Asymmetric Demand Information in Uniform and Discriminatory Call Auctions: An experimental Analysis Motivated by Electricity Markets. *Journal of Regulatory Economics*, 23 (2): 125 – 144.

Abreu, D., Rubinstein, A., 1988. The Structure of Nash Equilibrium in Repeated Games with Finite Automata. *Econometrica*, 56 (6): 1259-1281.

Allaz, B., 1992. Oligopoly, uncertainty and strategic forward transactions. *International Journal of Industrial Organization*, 10 (2): 297-308.

Allaz, B., Vila, J.-L., 1993. Cournot Competition, futures markets and efficiency. *Journal of Economic Theory*, 59 (1): 1-16.

Anderson, E. J., Philpott, A. B., 2002a. Using Supply Functions for Offering Generation into an Electricity Market. *Operations Research,* 50 (3): 477-489.

Anderson, E. J., Philpott, A. B., 2002b. Optimal Offer Construction in Electricity Markets. *Mathematics of Operations Research*, 27 (1): 82-100.

Anderson, E. J., Philpott, A. B., 2003. Estimation of Electricity Market Distribution Functions. *Annals of Operations Research*, 121: 21-32.

Angluin, D., 1987. Learning Regular Sets from Queries and Counterexamples. *Information and Computation*, 75: 87-106.

Arthur. W. B., 1991. Designing Economic Agents that Act like Human Agents: A Behavioral Approach to Bounded Rationality. *The American Economic Review*, 81 (2), Papers and Proceedings of the Hundred and Third Annual Meeting of the American Economic Association: 353-359.

Aumann, R. J., 1997. Rationality and Bounded Rationality. *Games and Economic Behavior*, 21: 2–14.

Banks, J. S., Sundaram, R. K., 1990. Repeated Games, Finite Automata, and Complexity. *Games and Economic Behavior*, 2: 97-117.

Barto A. G., Bradtke, S. J., Singh, S. P., 1995. Learning to act using real-time dynamic programming. *Artificial Intelligence*, 72: 81-138

Baxter, J., Tridgell, A., Weaver, L., 2000. Learning to Play Chess using Temporal Differences. *Machine Learning*, 40 (3): 243 - 263

Bellman, R., 1957. Dynamic programming, Princeton University Press

Ben-Porath, E., 1990. The Complexity of Computing a Best Response Automaton in Repeated Games with Mixed Strategies. *Games and Economic Behavior*, 2: 1-12.

Bernheim, D., 1984. Rationalizable Strategic Behavior. *Econometrica*,52: 1007-1028.

Bertsekas, D. P., Tsitsiklis, J. N., 1996. Neuro-Dynamic Programming. Athena Scientific, Belmont, Massachusetts.

Bertsekas, D. P., 2000. Dynamic Programming and Optimal Control. Vol. I and II, second edition. Athena Scientific, Belmont, Massachusetts

Bettis, R. A., Prahalad, C. K., 1995., The dominant logic: retrospective and extension. *Strategic Management Journal*, 16: 5-14.

Binmore, K. G., Shaked, A., Sutton, J., 1985. Testing Noncooperative Bargaining Theory: A Preliminary Study. *American Economic Review*, 75: 1178-1180.

Bodde, D. L., 1998. Strategic thinking about nuclear energy: implications of the emerging market structure in electricity generation. *Energy Policy*, 26 (12): 957-962.

Borenstein, S., Bushnell, J., Kahn, E., Stoft, S., 1995. Market power in California electricity markets. *Utilities Policy*, 5 (3-4): 219-236.

Borenstein, S., Bushnell, J., 1999. An Empirical Analysis of the Potential for Market Power in California's Electricity Industry. *The Journal of Industrial Economics*, XLVII (3): 285 – 323.

Borenstein, S., Bushnell, J., Knittel, C. R., 1999. Market Power in Electricity Markets: Beyond Concentration Measures. *The Energy Journal*, 20 (4): 65–88.

Borenstein, S., Bushnell, J., Wolak, F., 2002. Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market. *American Economic Review*, 92 (5): 1376 – 1405.

Bower, J., Bunn, D. W., 2000. A Model-Based Comparison of Pool and Bilateral Market Mechanisms for Electricity Trading. *The Energy Journal,* 21 (3): 1-29.

Bower, J., Bunn, D. W., 2001. Experimental Analysis of the Efficiency of uniform-Price v Discriminatory Auctions in the England and Wales Electricity Market. *Journal of Economic Dynamics and Control,* 25 (3-4): 561-592.

Bradtke, S. J., Barto, A. G., 1996. Linear Least-Squares Algorithms for Temporal Difference Learning. *Machine Learning*, 22: 33 - 57

Bunn, D. W., 1999. Electricity Markets are Different. Strategic price risk in wholesale power markets. Risk Books.

Bunn, D. W., 1997. Demand-side Participation in the Electricity Pool of England and Wales. Working Paper. Decision Technology Centre. Energy Markets Group. London Business School.

Bunn, D. W., Paschentis, S. N., 1983. Economic Dispatch of Electric Power By Stochastic Linear Programming. B. Lev (Editor), Energy Models and Studies. North-Holland publishing company.

Carmel, D., Markovitch, S., 1996. Learning Models of Intelligent Agents. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence.* Portland, Oregon: 62-67.

Carmel, D., Markovitch, S., 1999. Exploration Strategies for Model-based Learning in Multi-agent Systems. *Autonomous Agents and Multi-agent Systems*, 2: 141 – 172.

Cason, T. N., Friedman, D., 1997. Price formation in Single Call Markets. *Econometrica,* 65 (2): 311-345.

Cooper, R. W., DeJong, D. V., Forsythe, R., Ross, T. W., 1990. Selection Criteria in Coordination Games: Some Experimental Results. *The American Economic Review*, 80 (1): 218-233.

Cox, A. J., 1999. Mergers, Acquisitions, Divestures, and Applications for Market-Based Rates in a Deregulated Electric Utility Industry. *The Electricity Journal*, 12 (4): 27-36.

Crampes, C., Creti, A., 2001, October. Price Bids and Capacity Choice in Electricity Markets. Working paper, Institut D'Economie Industrielle, http://www.idei.asso.fr/.

Daughety, A. F, 1985. Reconsidering Cournot: The Cournot Equilibrium is Consistent. *The RAND Journal of Economics*, 16 (3): 368 – 379.

Day, C. J., Bunn, D. W., 2001. Divestiture of generation assets in the electricity pool of England and Wales: A computational approach to analysing market power. *Journal of Regulatory Economics*, 19 (2): 123-141.

Dayan, P., 1992. The Convergence of TD(λ) for General λ. *Machine Learning*, 8: 341-362

Electricity Association. 1999. The UK Electricity System, www.electricity.org.uk.

Electricity Association. 2000a. Electricity Companies in the United Kingdom – a brief chronology, www.electricity.org.uk.

Electricity Association. 2000b. List of Power Stations, www.electricity.org.uk.

Electricity Association. 2000c. Who owns whom in the UK electricity industry, www.electricity.org.uk.

Eliaz, K., 2003. Nash equilibrium when players account for the complexity of their forecasts. *Games and Economic Behavior*, 44: 286-310.

Elmaghraby, W., Oren, S. S., 1999. The Efficiency of Multi-Unit Electricity Auctions. *The Energy Journal*, 20 (4): 89-116.

Erev, I., Roth, A. E., 1998. Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, 88 (4): 848-881.

Exelby, M. J., Lucas, N. J. D., 1993. Competition in the UK Market for Electricity Generating Capacity, A Game Theory Analysis. *Energy Policy*, April: 348–354.

Feltovich, N., 2000. Reinforcement-Based Vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games. *Econometrica*, 68 (3): 605-641.

Ferris, M. C., Pang, J. S., 1997. Engineering and Economic Applications of Complementarity Problems. *SIAM Review*, 39 (4): 669-713.

Flavin, C., Lenssen, N., 1994. Reshaping the electricity power industry. *Energy Policy*, 22 (12): 1029–1044.

Franklin, S., Graesser, A., 1997. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*: 21-35.

Freund, Y., Kearns, M., Mansour, Y., Ron, D., Rubinfeld, R., Schapire, R. E., 1995. Efficient Algorithms for Learning to Play Repeated Games Against Computationally Bounded Adversaries. *In proceedings of the 36th IEEE Symposium on the Foundations of Computer Science*: 332-341.

Fudenberg, D., Tirole, J., 1991. Game Theory. The MIT Press, Cambridge, Massachusetts.

Fudenberg, D., Levine, D. K. 1998. The Theory of Learning in Games. The MIT Press, Cambridge, Massachusetts.

Garcia, A., Reitzes, J. D., Stacchetti, E., 2001. Strategic Pricing when Electricity is Storable. *Journal of Regulatory Economics*, 20 (3): 223-247.

Gilboa, I., 1988. The Complexity of Computing Best-Response Automata in Repeated Games. *Journal of Economic Theory*, 45: 342-352.

Gilboa, I., Zemel, E., 1989. Nash and Correlated Equilibria: Some Complexity Considerations. *Games and Economic Behavior*, 1: 80 – 93.

Gilboa, I., Samet, D., 1989. Bounded versus Unbounded Rationality: The Tyranny of the Weak. *Games and Economic Behaviour*, 1: 213-221.

Gold, E. M., 1978. Complexity of Automaton Identification from Given Data. *Information and Control*, 37: 302-320.

Green, R., Newbery, D., 1992. Competition in the British Electricity Spot Market. *Journal of Political Economy*, 100 (5): 929-953.

Green, R., 1996. Increasing Competition in the British Electricity Spot Market. *Journal of Industrial Economics*, XLIV (2): 201-216.

Green, R., 1999. The electricity contract market in England and Wales. *Journal of Industrial Economics*, XLVII (1): 107-124.

Harsanyi, J. C., Selten, R., 1988. A general theory of Equilibrium Selection in Games, Cambridge: MIT Press.

Henney, A., 1994. Energy markets and energy policies after the White Paper. *Energy Policy*, January: 5-116.

Herguera, I., 2000. Bilateral contracts and the spot market for electricity: some observations on the British and the NordPool experiences. *Utilities Policy*, 9 (2): 73-80.

Hobbs, B. F., 1986. Network Models of Spatial Oligopoly with an Application to deregulation of electricity generation. *Operations Research,* 34 (3), 395-409.

Hobbs, B. F., 2001, Linear Complementarity Models of Nash-Cournot Competition in Bilateral and POOLCO Power Markets. *IEEE Transactions on Power Systems*, 16 (2): 194-202.

Hopcroft, J. E., Ullman, J. D., 1979. Introduction to automata theory, Languages and Computation. Addison-Wesley, Mass.

Howard, R. A., 1971. Dynamic probabilistic systems. John Wiley.

Howard, R. A., 1960. Dynamic programming and Markov Processes. MIT Press.

Ishii, J., Yan, J., 2002, September. The "Make or Buy" Decision in U.S. Electricity Generation Investments. Working paper, CSEM WP 107, UCEI, www.ucei.org.

Joskow, P. L., 1997. Restructuring, Competition and Regulatory Reform in the U.S. Electricity Sector. *Journal Economic Perspectives*, 11 (3): 119-138.

Joskow, P. L., Kahn, E., 2002. A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000. *The Energy Journal*, 23 (4): 1 – 35.

Kaelbling L. P., Littman, M. L., Moore, A. W., 1996. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4: 237–285.

Kalai, E., Lehrer, E., 1993. Rational Learning Leads to Nash Equilibrium. *Econometrica*, 61 (5): 1019 – 1045.

Karp, R. M., Held, M., 1967. Finite-State Processes and Dynamic Programming. *SIAM Journal on Applied Mathematics*, 15 (3): 693-718.

Kaserman, D. L., Mayo, J. W., 1991. The measurement of vertical economies and the efficient structure of the electricity utility industry. *The Journal of Industrial Economics*: 483–502.

Kennedy, D., 1997. Merger in the English electricity industry. *Energy Policy*, 25 (4): 393–399.

Klemperer, P. D., Meyer, M. A., 1989. Supply Function Equilibria in Oligopoly under Uncertainty. *Econometrica,* 57 (6): 1243-1277.

Kreps, D. M., Scheinkman, J. A., 1983. Quantitative Precommitment and Bertrand Competition Yield Cournot Outcomes. *The Bell Journal of Economics*, 14 (2): 326-337.

Law, A. M., Kelton, W. D., 1991. Simulation Modelling and Analysis. Second Edition. McGraw-Hill.

Larsen, E. R., Bunn, D. W., 1999. Deregulation in electricity: understanding strategic and regulatory risk. *Journal of the Operational Research Society*, 50: 337-344.

Lomi, A., Larsen, E. R., 1996. Interacting Locally and Evolving Globally: a Computational Approach to the Dynamics of Organizational Populations. *Academy of Management Journal*, 39 (4): 1287-1321.

London Economics, 1999, October. Role Playing Simulations of The New Electricity Trading Arrangements, A Report to the Reta Programme. http://www.londecon.co.uk/

Lowrey, C., 1997. The Pool and forward contracts in the UK electricity supply industry. *Energy Policy*, 25 (4): 413–423.

Marcet, A., Nicolini, J. P., 2002, December. Recurrent Hyperinflations and Learning. Working paper, Universitat Pompeu Fabra, http://www.econ.upf.es/

Michalewicz, Z., 1992. *Genetic Algorithms + Data Structures = Evolution Programs*, 2nd ed. Berlin, Springer-Verlag.

Milgrom, P. R., 1987. Auction theory. In *Advances in Economic Theory, Fifth World Congress*, Truman Bewley, Ed. Cambridge University Press: 1-32.

Moore, A. W., Atkeson, C. G., 1993. Prioritized Sweeping: Reinforcement Learning With Less Data and Less Real Time. *Machine Learning*, 13: 103 – 130.

Mor, Y., Goldman, C. V., Rosenschein, J. S., 1996. Learn Your Opponent's Strategy (in Polynomial Time)!. G. Weiss and S. Sen (eds.). Lecture Notes in Artificial Intelligence, Vol. 1042, Springer Verlag.

Newbery, D. M., 1998. Freer electricity markets in the UK: a progress report. *Energy Policy*. 26 (10): 743-749.

Nicolaisen J., Petrov, V., Tesfatsion, L., 2001. Market Power and Efficiency in a Computational Electricity Market with Discriminatory Double-Auction Pricing. *IEEE Transactions on Evolutionary Computation*, 5 (5): 504- 523.

Ofgem, 1999, July. The New Electricity Trading Arrangements. http://www.ofgem.gov.uk/

Ofgem/DTI, 1999, October. The New Electricity Trading Arrangements, Ofgem/DTI Conclusions Document. http://www.ofgem.gov.uk/.

Ofgem, 2000, March. Settlement Administration Agent User Requirements Specification. http://www.ofgem.gov.uk/.

Ofgem, 2000, April. NGC Incentives Under Neta. http://www.ofgem.gov.uk/.

Papadimitriou, C. H., 1992. On Players with a Bounded Number of States. *Games and Economic Behavior*, 4: 122 – 131.

Pearce, D., 1984. Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica*, 52: 1029-1050.

Piccione, M., 1992. Finite Automata Equilibria with Discounting. *Journal of Economic Theory*: 56, 180-193.

Pineau, P.-O., Murto, P., 2003. An Oligopolistic Investment Model of the Finnish Electricity Market. *Annals of Operations Research*: 121, 123-148.

Power UK, 2002. News and Analysis of the UK Power Industry, 97, March.

Puller, S. L., 2002, August. Pricing and Firm Conduct in California's Deregulated Electricity Market. Working Paper, http://econweb.tamu.edu/puller/.

Ramos, A., Ventosa, M., Rivier, M., 1998. Modelling competition in electric energy markets by equilibrium constraints. *Utilities Policy*, 7: 233-242.

Rassenti, S. J., Smith, V. L., Wilson, B. J., 2003. Discriminatory Price Auctions in Electricity Markets: Low Volatility and the Expense of High Price Levels. *Journal of Regulatory Economics*, 23 (2): 109 – 123.

Ron, D., Rubinfeld, R., 1997. Exactly Learning Automata with Small Cover Time. *Machine Learning*: 27, 69-96.

Ross, S., 1983. Introduction to Stochastic Dynamic Programming. Academic Press.

Roth, A. E., Erev, I., 1995. Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term. *Games and Economic Behavior*, 8: 164-212.

Roth, A. E., 2002. The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. *Econometrica*, 70 (4): 1341 – 1378.

Rothkopf, M. H., 1999. Daily Repetition: A Neglected Factor in the Analysis of Electricity Auctions. *The Electricity Journal*, 11: 61-70.

Rubinstein, A., 1986. Finite Automata Play the Repeated Prisoner's Dilemma. *Journal of Economic Theory*, 39: 83-96.

Rubinstein, A. 1998. Modeling Bounded Rationality. MIT Press.

Rudkevitch, A., Duckworth, M., Rosen, R., 1998. Modelling Electricity Pricing in a Deregulated Generation Industry: The Potential for Oligopoly Pricing in a Pool. *The Energy Journal*, 19(3): 19-48.

Salama, A., 1997. Privatization and culture change: British Nuclear Fuels case study. *Energy Policy*, 25 (3): 293-304.

Samuel, A. L., 1959. The history of heuristic and alpha-beta search enhancements in practice. *IEEE Transactions on Pattern Analysis and Machine Learning*, 11 (11): 1203-1212

Samuelson, L., 1997. Evolutionary Games and Equilibrium Selection. The MIT Press.

Sarin, R., Vahid, F., 2001. Predicting how people play games: A Simple Dynamic Model of Choice. *Games and Economic Behavior*, 34: 104-122.

Schreider, Y. A., 1964. Automata and the problem of dynamic programming. *Problems of Cybernetics*: 5: 33-58.

Selten, R., 1975. Reexamination of the perfectness concept for equilibrium points in extensive-form games. *International Journal of Game Theory*, 4: 25-55.

Simon, H. A., 1972. Theories of bounded rationality, in Macguire and Radner (Eds.), Decision and Organisation, North-Holland.

Singh, S. P., 1994. Learning to solve markovian decision processes. Doctor of Philosophy Dissertation, Department of Computer Science, University of Massachusetts.

Singh, S., Jaakkola, T., Szespesvári, C., 2000. Convergence Results for Single-Step On-line Reinforcement Learning Algorithm. *Machine Learning*, 38 (3): 287 – 308.

Skantze, P., Visudhiphan, P., Ilic, M., 2000, November. Valuation of Generation Assets with Unit Commitment Constraints under Uncertain Fuel Prices. Working Paper MIT_EL 00-006.

Staiger, R. W., Wolak, F. A., 1992. Collusive Pricing with Capacity Constraints in the Presence of Demand Uncertainty. *The RAND Journal of Economics*, 23 (2): 203-220.

Stirling, A., 1994. Diversity and ignorance in the electricity supply investment. *Energy Policy*, March: 195–201.

Sutton, R., 1988. Learning to Predict by the Methods of Temporal Differences. *Machine Learning*, 3: 9-44.

Sutton, R. S., Barto, A. G., 1998. Reinforcement Learning: An Introduction. MIT Press.

Sweetser, A., 1998. Measuring a Dominant Firm's Market Power in a Restructured Electricity Market, a Case Study of Colorado. *Utilities Policy*, 7: 243-257.

Takriti, S., Krasenbrink, B., Wu, L. S.-Y., 2000. Incorporating Fuel Constraints and Electricity Spot Prices into The Stochastic Unit Commitment Problem. *Operations Research*: 48 (2): 268-280.

Tesauro, G., 1992. Practical Issues in Temporal Difference Learning. *Machine Learning*, 8: 257-277.

Tesauro, G., 1994. TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play. *Neural Computation*, 6: 215-219.

Tsitsiklis, J. N., Roy, B. V., 1997. An Analysis of Temporal-Difference Learning with Function Approximation. *IEEE Transactions on Automatic Control*, 42: 674 – 690.

van Damme, E., 1991. Stability and Perfection of Nash Equilibria. Second edition, Springer-Verlag.

van Huyck, J. B., Battalio, R. C., Beil, R. O., 1990. Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure. *The American Economic Review*, 80 (1): 234-248.

Visudhiphan, P., Skantze, P., Ilic, M., 2001, July. Dynamic Investment in Electricity Markets and Its Impact on System Reliability. Working paper MIT_EL 01-012WP.

von der Fehr, N.-H. M., Harbord, D., 1993. Spot Market Competition in the UK Electricity Industry. *The Economic Journal*, 103 (418): 531-546.

von Neumann, J., Morgenstern, O., 1953. Theory of Games and Economic Behavior. Third Edition. Princeton: Princeton University Press.

Ward, D. P., Murray, A. T., Phinn, S. R., 2000. A stochastic constrained cellular model of urban growth. *Computers, Environment and Urban Systems*, 24: 539-558.

Watkins, C. J. C. H., Dayan, P., 1992. Technical Note, Q-Learning. *Machine Learning*, 8: 279 – 292.

Weber, B., 1995. Bypass Trading and Market Quality in Electronic Security Exchanges. *Journal of Organizational Computing*, 5 (3): 327-353.

Wei, J.-Y., Smeers, Y., 1999. Spatial Oligopolistic Electricity Models With Cournot Generators and Regulated Transmission Prices. *Operations Research*, 47 (1): 102-112.

Weibull, J.W., 1995. Evolutionary Game Theory. The MIT Press.

Weiss, G., 1995. Adaptation and Learning in Multi-Agent Systems: Some Remarks and a Bibliography. *Adaptation and Learning in Multiagent Systems*, G. Weiss and S. Sen, Ed. Springer: 1-21.

Whitley, D., Kauth, J., 1988. Genitor: A Different Genetic Algorithm. Working paper, Technical Report CS-88-101, Department of Computer Science Colorado State University.

Wilson, R., 1987. Game-Theoretic Analysis of Trading Processes. In *Advances in Economic Theory: Fifth World Congress*, Truman Bewley, Ed. Cambridge University Press: 33-70.

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

Wolak, F. A., Patrick, R. H. 1997, February. The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market. Working Paper PWP-047, www.ucei.berkeley.edu/ucei.

Wolfram, C. D., 1998. Strategic Bidding in a Multiunit Auction: an Empirical Analysis of Bids to Supply Electricity in England and Wales. *RAND Journal of Economics*, 29 (4): 703 – 725.

APPENDICES

TABLE A.1: Generation Database

Producer	StationID	Capacity (MW)	Marginal Cost (£/MWh)
AES	AESB_	250	8.68
AES	DRAXX	645	12.60
AES	DRAXX	645	12.60
AES	DRAXX	645	12.60
AES	DRAXX	645	12.60
AES	DRAXX	645	12.60
AES	DRAXX	645	12.60
AES	DRAXX	75	69.00
AES	Fifoots Point	405	13.00
Barking	BARK_	600	14.69
Barking	BARK_	400	14.69
British Energy	Heym	590	3.00
British Energy	Heym	570	3.00
British Energy	Heym	660	3.00
British Energy	Heym	660	3.00
British Energy	HINB_	660	3.00
British Energy	HINB_	660	3.00
British Energy	HRTL	610	3.00
British Energy	HRTL	580	3.00
British Energy	EGGB_	490	12.09
British Energy	EGGB_	480	12.09
British Energy	EGGB_	505	12.09
British Energy	EGGB_	505	12.09
British Energy	SIZEA	1200	3.00
Derwent Cogen	DERW_	235	14.77
EdF	SUTB_	801	6.59

EdF	FREX_	1988	5.00
Edison	ROOS_	229	14.62
Edison	FERR_	490	12.56
Edison	FERR_	490	12.56
Edison	FERR_	490	12.56
Edison	FERR_	490	12.56
Edison	FIDL_	486	12.76
Edison	FIDL_	485	12.76
Edison	FIDL_	451	12.76
Edison	FIDL_	507	12.76
Edison	FERR_	34	42.00
Edison	FIDL_	34	46.00
Edison	DINO_	288	20.00
Edison	DINO_	288	20.00
Edison	DINO_	288	20.00
Edison	DINO_	288	20.00
Edison	DINO_	288	20.00
Edison	DINO_	288	20.00
Edison	FFES_	90	20.00
Edison	FFES_	90	20.00
Edison	FFES_	90	20.00
Edison	FFES_	90	20.00
Entergy	SCCL_	400	10.91
Entergy	SCCL_	400	10.91
Entergy	SCCL_	400	10.91
IPP1	REDD_	27	74.00
IPP2	FAWN	150	7.00
IPP3	FELL_	168	14.99
IPP4	SEAB_	812	13.13
IPP5	TESI_	937	8.79
IPP5	TESI_	938	8.79
IVO	BRGG_	270	14.42
Magnox	BRWE_	296	3.00
Magnox	OLDS_	460	3.00
Magnox	SIZEA	430	3.00
Magnox	WYLF_	1100	3.00
Medway Power	MEDP_	700	14.77
National Power	DIDC_	660	14.35
National Power	DIDC_	660	14.35
National Power	LBAR	660	14.35
National Power	DIDC_	490	12.10
National Power	DIDC_	450	12.10
National Power	DIDC_	490	12.10
National Power	DIDC_	490	12.10
National Power	ABTHB	485	12.10

National Power	ABTHB	485	12.10
National Power	ABTHB	485	12.10
National Power	BLYTB	313	16.18
National Power	BLYTB	313	16.18
National Power	TILBB	340	13.05
National Power	TILBB	336	13.05
National Power	ABTHB	34	50.00
National Power	COWE_	70	94.34
National Power	COWE_	70	54.25
National Power	DIDC_	50	30.00
National Power	DIDC_	50	50.00
National Power	FAWL_	34	30.00
National Power	LITTD	70	40.00
National Power	TILBB	34	28.00
National Power	FAWL_	484	30.00
National Power	LITTD	685	29.00
NRG	EECL_	408	13.05
NRG	KILLN	650	14.54
PowerGen	CDCL_	525	13.28
PowerGen	CORB_	406	13.28
PowerGen	KILLP	450	13.28
PowerGen	KILLP	450	13.28
PowerGen	RYEH_	719	13.28
PowerGen	KINO_	485	12.37
PowerGen	KINO_	485	12.37
PowerGen	KINO_	485	12.37
PowerGen	COTT_	497	13.25
PowerGen	COTT_	497	13.25
PowerGen	COTT_	517	13.25
PowerGen	COTT_	497	13.25
PowerGen	RATS_	500	12.44
PowerGen	RATS_	500	12.44
PowerGen	RATS_	500	12.44
PowerGen	RATS_	500	12.44
PowerGen	GRAI_	58	60.00
PowerGen	INDQ_	140	28.00
PowerGen	KINO_	34	50.00
PowerGen	RATS_	34	44.00
PowerGen	TAYL_	68	72.00
PowerGen	TAYL_	64	75.00
PowerGen	GRAI_	675	29.00
PowerGen	GRAI_	675	29.00
Rocksavage Power	ROCK_	760	13.36
S&S	KEAD_	760	14.32
S&S	HYDEX	350	5.00

Scottish Power	SPEX_	655	5.00
South Humber Power	SHBA_	792	13.36
South Humber Power	SHBA_	528	13.36
TXU	KLYNA	380	10.40
TXU	PETEM	405	10.40
TXU	RUG	488	13.86
TXU	RUG	488	13.86
TXU	WEB	483	14.22
TXU	WEB	503	14.22
TXU	WEB	503	14.22
TXU	WEB	483	14.22
TXU	DRK	333	18.27
TXU	DRK	333	18.27
TXU	DRK	333	18.27
TXU	RUG	50	50.00
TXU	WEB	40	35.00
TXU	HMAR	189	15.87
TXU	HMAR	189	15.87
TXU	HMAR	189	15.87
TXU	HMAR	189	15.87
TXU	HMAR	189	15.87

TABLE A.2: : Power Exchange Prices (£/MWh), average of the last 10 days, for each one of the six strategies in each one of the six scenarios simulated (twelve simulations)

		Strategies					
		ST1	ST2	ST3	ST4	ST5	ST6
Simulation	One	21.8	22.4	19.2	22.8	22.8	30.5
	Two	17.1	18.7	20.6	18.4	20.6	19.2
	Three	14.1	14.5	16.7	15.1	17.5	17.1
	Four	13.9	14.1	14.2	14.5	14.1	14.3
	Five	14.3	14.4	14.3	14.0	14.4	14.6
	Six	14.4	14.5	14.4	14.7	14.6	15.6
	Seven	19.0	20.2	21.8	24.0	35.3	28.0
	Eight	18.0	17.2	19.7	21.2	20.0	24.6
	Nine	16.3	17.0	16.6	15.7	18.1	18.1
	Ten	15.9	14.7	20.4	16.6	28.4	25.7
	Eleven	14.3	14.4	16.0	16.7	15.5	16.3
	Twelve	14.4	14.4	14.4	14.4	15.1	14.6

TABLE A.3: System Sell Prices (£/MWh), average of the last 10 days, for each one of the six strategies in each one of the six scenarios simulated (twelve simulations)

		Strategies					
		ST1	ST2	ST3	ST4	ST5	ST6
Simulation	One	7.0	6.1	8.4	6.9	6.4	7.2
	Two	7.7	5.7	7.7	6.5	6.1	8.2
	Three	6.7	8.9	7.8	6.5	8.5	6.2
	Four	8.4	8.4	5.7	8.8	8.4	9.3
	Five	4.3	6.9	8.0	9.9	7.9	7.2
	Six	7.9	6.7	6.0	10.4	8.0	8.7
	Seven	7.9	6.3	9.8	9.6	7.7	9.1
	Eight	6.5	5.1	7.5	7.5	5.2	7.0
	Nine	4.9	7.4	5.0	8.9	7.6	5.6
	Ten	6.2	7.8	5.3	8.2	6.5	7.9
	Eleven	8.5	7.7	7.8	6.9	5.9	6.5
	Twelve	6.8	7.0	8.8	6.7	6.3	6.9

TABLE A.4: System Buy Prices (£/MWh), average of the last 10 days, for each one of the six strategies in each one of the six scenarios simulated (twelve simulations)

		Strategies					
		ST1	ST2	ST3	ST4	ST5	ST6
Simulation	One	31.5	42.2	25.7	33.9	45.4	48.8
	Two	28.8	29.7	36.3	30.3	26.4	27.7
	Three	21.6	23.0	30.1	20.4	35.4	23.5
	Four	19.3	22.0	19.1	17.4	23.6	24.3
	Five	21.2	25.9	17.4	19.7	17.4	20.9
	Six	28.8	16.3	15.9	18.6	17.8	17.0
	Seven	35.1	36.5	41.7	46.2	69.3	45.9
	Eight	26.1	24.7	40.2	31.8	32.2	34.9
	Nine	25.9	22.2	30.0	27.3	28.3	27.0
	Ten	28.1	21.7	30.1	30.4	43.7	40.0
	Eleven	24.0	24.4	25.5	25.9	30.1	24.3
	Twelve	23.1	22.0	25.2	24.6	23.9	20.2

TABLE A.5: AES' s profits (£/hour), average of the last 10 days, for each one of the six strategies in each one of the six scenarios simulated (twelve simulations)

		Strategies					
		ST1	ST2	ST3	ST4	ST5	ST6
Simulation	One	21342	29809	20465	23387	11786	50133
	Two	8871	8117	18958	13991	20872	13716
	Three	4969	5280	9647	4489	10450	4548
	Four	4697	4247	4262	4043	4517	4548
	Five	6585	5632	3250	3627	4920	5608
	Six	4498	4695	5948	2905	1965	5854
	Seven	17616	21716	29593	26298	68445	25294
	Eight	15987	13015	14022	21627	17645	26420
	Nine	9782	7783	10659	8724	16181	11207
	Ten	10519	6000	19679	9584	41646	25758
	Eleven	7190	5484	8055	9259	8608	8996
	Twelve	6261	5472	5283	6318	5375	6097

TABLE A.6: **BE' s profits (£/hour),** average of the last 10 days, for each one of the six strategies in each one of the six scenarios simulated (twelve simulations)

		Strategies					
		ST1	ST2	ST3	ST4	ST5	ST6
	One	104554	74738	80757	101959	85079	128300
	Two	74687	74021	96000	72482	56638	58847
	Three	32292	34480	45565	52878	78948	51957
	Four	43848	26669	33296	35740	65664	51957
Simulation	Five	47361	35443	33356	37040	46429	56716
	Six	29523	52630	53589	56962	62458	67605
	Seven	63150	64746	65095	84040	115194	90837
	Eight	60463	55585	52367	54282	50087	71383
	Nine	52738	44544	51400	38729	52436	57572
	Ten	37679	36889	62500	41021	101374	84526
	Eleven	44875	46060	45568	40979	39159	48663
	Twelve	43366	43208	33094	45702	38931	46384