

COMMERCIAL RISK MANAGEMENT IN THE ELECTRICITY SUPPLY INDUSTRY

David Yaw Adjepon-Yamoah



Thesis submitted for the degree of
Doctor of Philosophy
University of Edinburgh
January 2002



Abstract

The introduction of the New Electricity Trading Arrangements (NETA) in the UK from 31st March 2001 changed the nature of electricity trading from a centrally traded marginal pricing mechanism, known as the Pool, to a series of bilateral markets. This changed the nature of the risks facing market participants notably, higher price volatility, no single reference price on which to base long-term contracts and potentially punitive imbalance charges for participants whose commitments were not met.

It would appear that Suppliers will be the major casualties of the changes because their function under the previous system was primarily the billing and metering of customers whereas the introduction of NETA means that they must submit exact information about their demand requirements and contracted position to the System Operator, 4 hours in advance of each half-hour. Any shortfalls between expected and metered volume will attract prohibitive imbalance charges. Accurate forecasting is essential.

This thesis describes effective forecasting methodologies that can be used by Suppliers to forecast their half-hourly demand at Bulk Supply Level. Artificial Neural Networks were selected as the most effective forecasting approach from an examination of the various short-term forecasting methods and by analytical comparison with Multiple Linear Regressions.

The Artificial Neural Network (ANN) was optimised and configured to suit NETA's requirements. The optimised ANN was used to forecast according to the time-of-day, and the day-of-week, to determine the most accurate configuration. Both models had their strengths, with the parallel method being more accurate but the serial or linear, easier to implement. The methods designed achieved better financial results when compared with the current industrial standard.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof H.W. Whittington for his ability to inspire initiative and motivation, wise advice, support and encouragement during the time spent on this project. Also to Dr Robin Wallace for his help in proof reading this thesis. I am very grateful to Dr Peter Edwards for his patience and technical support in understanding and implementing Artificial Neural Networks.

I am very grateful to the help I received from all the past and present members of staff at Scottish Hydro Electric plc, particularly Dr Alan Birch, Dr Sandy Wito (currently with IPA consulting) and Mr John Connelly (retired).

I would like to thank all past and present members of the Energy Systems Group especially Dr C. L Silvertown, for their moral support. I am very grateful to the technical staff and the secretaries in the Department of Electrical & Electronic Engineering at University of Edinburgh for their help during the course of this study.

Also I would like to acknowledge all those who have helped and supported me both inside and outside the university arena over the many years of my education most notably, the Ogunro family.

Finally, I am greatly indebted to my parents, Kenneth and Beatrice, my siblings, Kenneth and Charlotte, for their love, patience and financial support, which made the completion of this research possible.

To my parents, Kenneth and Beatrice

Table of Contents

Abstract	i
Declaration of originality	ii
Acknowledgements.....	iii
Table of Contents	v
1. Introduction.....	1
1.1 Introduction	1
1.2 Project aims and objectives	2
1.3 Statement of thesis	3
1.4 Thesis outline	3
2. The UK Electricity Supply Industry.....	6
2.1 Overview	6
2.2 Introduction to the Electricity Supply Industry (ESI).....	6
2.2.1 Generation	6
2.2.2 Transmission.....	7
2.2.3 Distribution.....	8
2.2.4 Supply	8
2.3 Background to the changes in the Electricity Supply Industry	10
2.3.1 The early years.....	10
2.3.2 Pre-Privatisation.....	10
2.3.3 Privatisation.....	12
2.4 The Market structure – Pool.....	13
2.4.1 Competition in Generation	14
2.4.1.1 The experience so far in England and Wales.....	15
2.4.1.2 The experience so far in Scotland	17
2.4.2 Competition in Transmission.....	17
2.4.3 Competition in Supply	18
2.4.3.1 The experience so far in England and Wales.....	19
2.5 The Present situation.....	20
2.5.1 Power market games and strategy through the Pool.....	21
2.5.2 Power games and strategy through the contracts market.....	23
2.5.3 Power games and strategy through the capacity payment mechanism	24
2.5.4 Discussion.....	24
2.6 Summary	25
3. The New Trading Arrangements (NETA).....	26
3.1 Overview	26
3.2 NETA.....	26
3.2.1 Long-term markets.....	28
3.2.1.1 Forwards.....	28
3.2.1.2 Futures.....	28
3.2.2 Short-term bilateral market.....	29
3.2.3 Balancing Arrangements.....	30
3.2.3.1 Balancing points on the network.....	30

3.2.3.2	Balancing Mechanism (BM) operation	31
3.2.4	Cash-Out and Settlement Arrangements	32
3.2.4.1	Imbalance Volumes	32
3.2.5	Regulation	32
3.2.5.1	Governance	33
3.3	Scottish Trading Arrangements: British Electricity Trading and Transmission Arrangements (BETTA)	33
3.4	NETA's Implications on the ESI	34
3.4.1	The 'NETA effect'	35
3.4.1.1	Bilateral versus Marginal pricing	35
3.4.1.2	Abolition of capacity payments	35
3.4.1.3	Demand-side participation	36
3.4.1.4	Possible developments of financial markets	36
3.4.2	Implications of NETA on the various groups	36
3.4.2.1	Large portfolio generators	36
3.4.2.2	Nuclear	37
3.4.2.3	Independent Power Producers (IPPs)	37
3.4.2.4	CHP, Renewables, Embedded Generation	37
3.4.2.5	Suppliers	38
3.4.3	NETA's Losers	39
3.5	Summary	39
4.	Risk Management within the ESI	41
4.1	Overview	41
4.2	Definition of 'Risk'	41
4.2.1	Risk Management	42
4.3	General guidelines for the implementation of a risk management strategy	42
4.3.1	The risk of risk management	43
4.4	Quantifying Risk exposure	44
4.4.1	Statistical	44
4.4.2	Scenario Analysis	45
4.5	Risk management methods – hedging	45
4.5.1	Hedging strategies	46
4.6	Overview of risks in the Electricity Supply Industry	47
4.6.1	Credit risk	48
4.6.2	Operational risk	49
4.6.3	Basis risk	49
4.6.4	Price risk	50
4.6.5	Volumetric Risk	51
4.6.6	Summary of possible evolution of risks under NETA	53
4.7	Risk measurement and management in the ESI	54
4.7.1	Determining the level of exposure to Pool price risk	54
4.7.1.1	Inadequacy under NETA	55
4.7.2	Determining the level of volumetric risk	55
4.7.2.1	Larger customers	55
4.7.2.2	Smaller customers	57
4.7.2.3	Application of volumetric risk management tools under NETA	58

4.8	Risk management tools	58
4.8.1	The CFD Market	58
4.8.1.1	Stacking contracts	59
4.8.1.2	Time-dependent contracts	59
4.8.2	The Electricity Forward Agreement (EFA).....	60
4.8.3	Swing options	62
4.8.3.1	Price-driven swing options	62
4.8.3.2	Demand-driven swing options	63
4.8.4	Weather derivatives.....	63
4.8.5	Demand Side Management.....	65
4.8.6	Portfolio management	68
4.8.6.1	Value-at-risk	68
4.8.6.2	Profit-at-risk.....	69
4.8.7	Metering	69
4.9	The future of risk management under NETA.....	70
4.9.1	Price risk management	70
4.9.2	Volumetric risk management.....	71
4.10	Summary	72
5.	Electricity demand forecasting	73
5.1	Overview	73
5.2	Customer demand	73
5.2.1	Industrial users.....	73
5.2.2	Domestic/Commercial users.....	74
5.3	Fundamental demand drivers	74
5.3.1	Economic factors	75
5.3.2	Time factors.....	75
5.3.2.1	Seasonality.....	75
5.3.2.2	Time-of-use.....	77
5.3.2.3	Special days	78
5.3.2	Random factors.....	78
5.3.3	Weather effects	79
5.3.3.1	Space heating	80
5.3.3.2	Water heating.....	82
5.3.3.3	Refrigeration and freezers	83
5.3.3.4	Others	83
5.4	Forecasting	83
5.4.1	Forecasting time-frame	84
5.5	Forecasting methods	85
5.5.1	Regression methods	85
5.5.1.1	Multiple Linear Regression (MLR)	85
5.5.1.2	Polynomial and trigonometric regressions	88
5.5.1.3	Non-parametric regression	89
5.5.2	Time series methods.....	91
5.5.2.1	Moving Average (MA) process	91
5.5.2.2	Autoregressive (AR) process.....	91
5.5.2.3	Autoregressive Moving-Average (ARMA) process	92

5.5.2.4	Autoregressive integrated moving average (ARIMA) process	92
5.5.2.5	Transfer Function (TF) modelling	92
5.5.3	End-User models.....	93
5.5.4	Artificial Intelligence methods	94
5.5.4.1	Knowledge-based systems.....	94
5.5.4.2	Fuzzy Logic	95
5.5.4.3	Artificial neural networks.....	96
5.5.4.4	Hybrid systems.....	99
5.6	Comparison of the forecasting methods	99
5.6.1	Time series models	99
5.6.2	Regression methods	100
5.6.3	Artificial intelligence	100
5.6.4	Model choice	100
5.7	Summary	101
6.	Data issues and selection of forecast inputs.....	103
6.1	Overview	103
6.2	Data Selection.....	103
6.2.1	Demand data	103
6.2.2	Weather data.....	104
6.2.3	Time horizon.....	104
6.2.4	Software choice.....	104
6.2.4.1	Artificial Neural Network.....	105
6.3	Characteristics of the load.....	106
6.3.1	Radio Teleswitching	106
6.3.2	Dynamic load control.....	107
6.3.3	Customer heating tariffs	107
6.3.4.1	Real-time example of the effect of switch ON, on BSP demand	109
6.3.5	Complexities of forecasting teleswitching demand	110
6.4	Forecast configuration	111
6.4.1	Parallel forecasts	111
6.4.2	Linear forecasting	112
6.4.3	Choice of the most appropriate configuration	113
6.5	Choice of appropriate inputs for the forecast model	113
6.5.1	Correlation analysis.....	114
6.5.2	Statistical analysis on weather variables	114
6.5.2.1	Correlation analysis on weather variables.....	117
6.5.3	Correlation analysis on week day effect	118
6.5.4	Correlation analysis on the effect of previous half-hourly demand.....	120
6.5.5	Model Inputs.....	122
6.6	Summary	123
7.	Model construction and comparison	124
7.1	Overview	124
7.2	Forecast evaluation	124
7.3	Multiple Linear Regression.....	125
7.3.1	Modelling issues	125
7.3.1.1	Dummy variables	125

7.3.1.2	Multicollinearity	126
7.3.1.3	Heteroscedasticity	127
7.3.1.4	Autocorrelation	127
7.3.2	Model construction	127
7.3.3	Model validation	129
7.3.3.1	Test statistics	129
7.3.3.2	Test for autocorrelation	130
7.4	Artificial Neural Network	131
7.4.1	ANN data preparation	132
7.4.1.1	Encoding categorical information	132
7.4.1.2	Input data pre-processing	133
7.4.2	Training the network	135
7.4.3	Designing a well trained model: Cross validation	137
7.4.4	Network architecture	139
7.5	Naïve forecasting	140
7.6	Results and discussion	140
7.7	Conclusions	144
8.	Neural network forecast optimisation	146
8.1	Overview	146
8.2	Assessing an alternative encoding method	146
8.3	Training period	148
8.3.1	Realistic input re-evaluation	149
8.3.2	Optimising the length of the training window	151
8.4	Complete forecasting model	152
8.5	Alternative forecasting approach	154
8.5.1	Network selection: Comparison of the committee and holdout methods	156
8.5.2	Linear versus parallel forecasting	157
8.6	Results and discussion	158
8.6.1	Monday	158
8.6.2	Tuesday	159
8.6.3	Wednesday	160
8.6.4	Thursday	160
8.6.5	Friday	161
8.6.6	Saturday	162
8.6.7	Sunday	163
8.7	Further discussion	163
8.7.1	Accuracy of models	163
8.7.1.1	Accuracy of the day of week predictions	163
8.7.1.2	Accuracy of time of day predictions	164
8.7.2	Predictability of the errors	165
8.7.2.1	Predictability of the demand shape	166
8.7.2.2	Predictability of the error on each half-hour of the day	166
8.7.3	A simple risk management strategy	167
8.7.3.1	Risk management strategy using the parallel method	167
8.7.3.2	Risk management strategy using the linear method	168
8.8	Case study	168

8.9 Conclusions 169

8.10 Summary 170

9. Conclusion and Summary 172

9.1 Overview 172

9.2 Implications of NETA 172

9.3 Risk and its management under NETA 172

9.4 Forecasting Approaches 173

9.5 Data issues 174

9.6 Most effective approach 175

9.7 Optimising the ANN 175

9.8 Parallel versus linear forecasting 176

9.9 Application of approach 176

9.10 Suggestions for future work 177

9.10.1 Managing the volumetric risk 177

9.10.2 Predicting and managing the risk of forecast error 177

9.10.3 Application of models to Generator forecasting 178

9.10.4 Hybrid forecasting 178

9.11 Final conclusions 178

References 179

Appendix 1 Published papers184

Appendix 2 Prices195

1. Introduction

1.1 Introduction

Following the successful privatisation of British Gas and British Telecommunications in the mid-eighties, the government under Margaret Thatcher decided to privatise the Electricity Supply Industry (ESI), and this was successfully completed on 1st April 1990. Its primary aim was to deliver lower prices to customers and to encourage greater efficiency from competition. From this period onwards the emphasis of the industry would change from the technical considerations to economic efficiency.

The experiences over the first seven years after privatisation were mixed. The price of electricity should have fallen in accordance with the reducing cost of generation and greater competition. Instead, prices went up in real-terms and it was difficult for new participants to enter the market in order to increase competition in the price setting process. It was suggested that the larger Generators had taken advantage of their market power to receive and maintain high revenues by;

- Setting the marginal price of electricity, System Marginal Price (SMP), artificially high in the certain knowledge that they had no genuine competition because all scheduled generators received SMP;
- During the periods when SMP was low, generators held back capacity to inflate the Capacity Payments (CP), the incentive they received for making capacity available. This was to make up the loss of revenue;
- Additionally, by increasing the price volatility through holding back capacity, they strengthened their position in long-term contract negotiations with any Supplier who wanted to guarantee stable income flows.

It was clear that the initial trading arrangements were not working as intended. Soon after Labour's electoral victory in 1997, they instructed the Electricity Regulator, Professor S. C. Littlechild to investigate reforming the electricity trading arrangements. A document was published in July 1998 with proposals to replace the marginal pricing system with a series of bilateral markets operating from the long-term up until close to real-time. The proposed trading arrangements were known as the New Electricity Trading Arrangements (NETA). Their main objective was to create more efficiency by financially rewarding generator flexibility and penalising Supplier wastages through the balancing market and imbalance charges respectively. Under NETA there was a financial incentive for Generators to move towards more flexible plant

and operate closer to the system margins. After several delays, NETA came into effect on 31st March 2001.

Suppliers' principal function under the Pool was to carry out the billing and metering functions of their customers. They bought electricity directly from the Pool or through contracts linked to Pool price and had no input in the price setting process. Suppliers have a high proportion of fixed costs and have a very thin margin for error or mismanagement. Under NETA Suppliers now have the responsibility for purchasing their long and short-term demand volumes bilaterally. Additionally, they are required to inform the system operator of their contracted volume and expected demand, 4 hours before the scheduled period. Any differences between their expectations and metered volumes attract punitive imbalance charges.

Until recently the system operator, the National Grid Company (NGC) has performed demand forecasts primarily for operational purposes and the majority of these were medium to long-term. Under NETA, both Suppliers and Generators require accurate forecasts to inform them of their demand expectations and efficient generator scheduling respectively. Effective short-term forecasting will become increasingly integral to the financial viability of market participants. This thesis reports on work that developed a short-term methodology for NETA's participants that incorporates the weather, the primary driver of demand, into its forecasts. The analysis focuses on predicting demand of a Supplier's 33kV Bulk Supply Point (BSP). Scottish Hydro-Electric (SHE) was the Supplier chosen. They had installed extensive demand-side management measures which distorted the demand and posed extra challenges to forecasting.

Artificial Neural Networks (ANN) were selected from an examination of the various short-term forecasting approaches available and this was compared analytically with Multiple Linear Regressions (MLR). The Neural Network was optimised and configured to meet NETA's requirements. The most effective configuration both in terms of accuracy and ease of implementation (whether to predict according to the time-of-day or day-of week) was investigated to provide a definitive answer as to the best configuration to use when predicting demand.

1.2 Project aims and objectives

This project had several distinct objectives:

1. To investigate NETA to determine its implications and the sources of risk to the various participants.
2. To examine risk measurement and management tools under the Pool to predict their evolution under NETA.

3. To analyse the key short-term demand forecasting approaches to determine their limitations so as to recommend the most appropriate ones for forecasting under NETA.
4. To perform analytical comparison of the most appropriate short-term methods.
5. To assess whether it is possible to predict accurately the load from a dataset contaminated by active demand-side management.
6. To optimise and predict the demand from the most appropriate model, comparing the strengths of intra and inter day forecasting.
7. To determine the effectiveness of the approach using real-life NETA balancing market prices.

1.3 Statement of thesis

The thesis of this study is twofold.

- It should be possible to forecast accurately at a Bulk Supply Point supplying a load contaminated by active demand-side management to suit NETA's requirements using an ANN with weather variables among the predictors.
- Predicting according to time-of-day should be more desirable than according to day-of-week because there should be less noise within the data set used to train the individual models.

1.4 Thesis outline

This thesis suggests an approach for forecasting short-term demand under NETA using Artificial Neural Networks, which guarantees market participants greater financial certainty than the present industrial standard.

Chapter 2 gives an introduction to the basic functions of the ESI, this is followed by a description of privatisation and the initial trading arrangements, the Pool. The limitations with the original market structure are discussed. Chapter 3 gives a detailed description of the New Trading Arrangements (NETA). The most probable effects of the changes on the ESI are discussed together with an analysis of its impact on the various classes of market participants. The chapter concludes by focussing on Suppliers because they face greatest financial insecurity under the proposed changes.

Chapter 4 examines the subject of risk management within the ESI with a view to evolving robust strategies to help insulate Suppliers from NETA's uncertainties. The definition of 'risk' is presented in the first section together with a general description of financial risk management.

Chapter 1. Introduction

The focus is then narrowed to risk management and measurement within the ESI and the inadequacy of the present methods under NETA. This is followed by suggestions for accurate short-term demand forecasting to help Suppliers eliminate their exposure to imbalance prices.

Chapter 5 starts by discussing demand and the factors influencing its behaviour notably weather, time, random events and economic factors. The various demand forecasting tools are discussed in the second half of this chapter and Artificial Neural Networks and Multiple Linear Regressions are chosen for comparison. This is because the MLR is easily implemented and is the most widely used. The ANN is the most popular Artificial Intelligence method and is widely acknowledged as the best method for representing a 'cause-effect' relationship.

In Chapter 6 the data issues involved in forecasting demand at BSP level are discussed. This is followed by an introduction and discussion of the teleswitching system used by SHE for demand-side management and its implications to short-term forecasting. The forecast configuration of the models is discussed next. This was to determine whether time-of-day or day-of-week forecasting was the most appropriate. It was decided to use time-of-day forecasting for the initial comparisons and this assumption was tested in Chapter 8. This was followed by a statistical analysis of the data to select the most relevant inputs for the forecasting model.

In Chapter 7 the forecast evaluation method used in the assessment is described followed by an outline of the issues considered in constructing a correctly specified Multiple Linear Regression forecasting model. Details are given of the model construction and validation prior to discussing the Artificial Neural Network data pre-processing, training and construction. The most commonly used methodology within the ESI for performing naïve forecasts is presented and is used to illustrate the superiority of mathematical forecasting models over intuitive methods. From the results and discussion of the comparisons, the ANN model emerges as the better approach because it is capable of representing the high degree of nonlinearity in the load.

The optimisation of the ANN is described in Chapter 8 and the inputs configured to meet realistic NETA conditions. The first step in the optimisation was to compare forecast accuracy of the binary encoding methodology with an alternative. In an attempt to shorten the training time whilst improving the accuracy, the length of the training window and forecast horizon were re-examined. Before building the complete model, the validity of the chosen input architecture was assessed by constructing a model for the first six hours of the day. This revealed some inherent flaws, the input variables were re-evaluated and a revised model was built followed by the construction of 48 half-hourly forecasting models. Comparing the performance of the completed parallel method with a linear equivalent, tested the assumption that parallel methods are superior to serial or linear methods. This chapter ends with a simulation using real NETA balancing mechanism prices to confirm the validity of the forecast methods developed.

Chapter 1. Introduction

Finally Chapter 8 draws the main arguments and findings of the thesis together to present a summary and suggestions for future work.

2. The UK Electricity Supply Industry

2.1 Overview

This chapter gives a brief description of the UK Electricity Supply Industry (ESI). An introduction to its basic functions is given in the first section. This is followed by a description of privatisation and initial trading arrangements. The limitations of the original market structure are discussed and changes explained.

2.2 Introduction to the Electricity Supply Industry (ESI)

The simplest function of the ESI is the conversion of fuel and primary energy to electricity and its transportation to customers. Electricity is expensive to store so supply and demand have to be matched continuously. The following section gives an introduction to the basic functions of the ESI.

The ESI can be divided into the following functions

- **Generation** - The production of electricity
- **Transmission** - High voltage bulk transportation of electricity
- **Distribution** - Transportation and delivery from high voltage Bulk Supply Points (BSP) to customers
- **Supply** - Wholesale purchasing and retailing to end users

2.2.1 Generation

Generation is the primary function of the ESI and it involves converting either hydrocarbon, hydraulic or atomic energy into electricity. In the past generation was carried out on a large scale to utilise the cost advantages of economies of scale. Over the years generation sets (gensets) have got bigger as technology has improved. The modern gensets have greatly improved efficiencies over older ones, for example the modern 660MW coal fired units at Drax, boasted efficiencies of 38% compared with less than 33% for the previous 120MW units¹. The introduction of Combined-Cycle Gas Turbine (CCGT) systems fired with natural gas, in the early 1990's, led to an improvement in efficiency to over 55% from about 38% available from the old style Open Cycle Gas Turbines.

Over the past decade, CCGT stations have emerged as the plant of preferred choice. This is because CCGT are compact and produce less emissions than other fossil fuel generation technologies. They require fewer operators and can be built and decommissioned relatively quickly at low cost. Another reason for CCGT's ascension has been the reduction in gas prices over the last decade from 22p/therm to about 9p/therm². This was due to the gas market deregulation and the expiry of long-term 'must-take' fuel contracts with British Gas.

The total UK generation capacity at the end of September 1999 was 72 GW. The generation mix over the last decade is illustrated in figure 2.1. This figure shows the fact that gas has overtaken coal as the main source of generation and the output from nuclear has increased due to the improved efficiency of the stations and the 'must take' status afforded nuclear as part of the Non Fossil Fuel Agreement (NFFO)

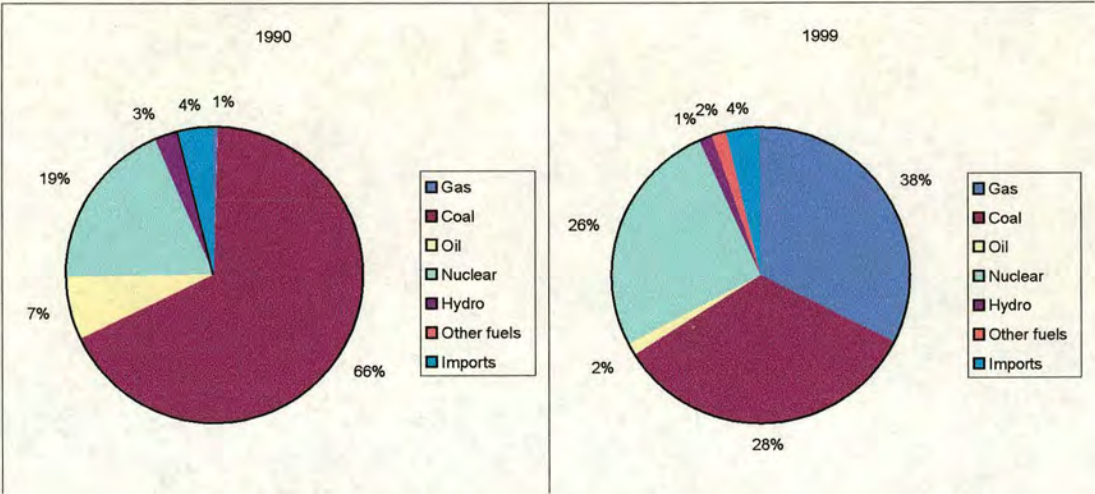


Figure 2 .1. Electricity Supplied by fuel type in 1990 and 1999 as percentages.

In the past generating stations tended to be located in areas of fuel availability. As such most of the coal stations commissioned in the 1950s and 60s were located in Northern England. The South-East has the highest concentration of CCGT generation because at present it is the location of the highest demand. Figure 2.2 shows a schematic diagram of the UK's electricity infrastructure.

2.2.2 Transmission

Transmission is the bulk transport of electricity at High Voltage (HV). The electricity is transported at HV because line losses are inversely proportional to the square of voltage. In England and Wales the network consists of 275kV and 400kV lines, whereas in Scotland the network uses 400/275/132kV, while 275kV and 132kV lines serve Northern Ireland.

Transmission is regarded as a natural monopoly because duplication would be uneconomical and lead to under use of the network. Control of the network is by the National Grid Company (NGC) in England and Wales, Scottish Power (SP) and Scottish Hydro-electric (SHE) in Scotland, and Northern Ireland Electricity (NIE) in Northern Ireland. Figure 2.2 shows all lines at or above 275kV on the transmission network in mainland Britain.

The network transmission operators have the following obligations to their customers

- To maintain the stability and security of the system
- Match supply and demand
- Cover line losses

The transmission operators do this by contracting plant to generate or be held in reserve in cases of emergency. The extra capacity is collectively known as ancillary services.

2.2.3 Distribution

Distribution involves the transport of electricity from Bulk Supply Point (BSP) to consumers through successively lower voltage circuits. The BSPs are defined as metered points at the step-down transformer, interfacing the HV transmission grid with the distribution network. In England and Wales, the distributors take power off the BSPs onto their 132kV networks. The voltage is stepped down to 33kV and 11kV and then finally to 400/230V, it is subsequently delivered to households and small customers. Some industrial and commercial customers take power at 11kV and 400kV. For example, National Power has a contract to supply an industrial gas manufacturer directly from its Eggborough coal fired station via a short 11kV link

Similar to transmission, the high cost of developing the network means distribution is considered a natural monopoly. There are 12 Regional Electricity Companies (REC) covering England and Wales. In Scotland and Northern Ireland, the situation is rather different; here the transmission companies provide the distribution. The RECs boundaries are shown in figure 2.2.

2.2.4 Supply

This involves the wholesale purchase of electricity via contracts in the wholesale market and retail to the end-user customers. In addition to the energy purchase Suppliers have to pay a charge known as 'use-of-system charges' for using the distribution and transmission networks. They are responsible for metering and billing the end-user. In England and Wales the supply function was initially performed by the RECs. Over time successive tranches of customer groups were opened to competition between Suppliers to drive prices down.



**Figure2.2. UK ESI showing the Gensets, transmission network the England and Wales
RECs and Scottish companies**

2.3 Background to the changes in the Electricity Supply Industry

The ESI has been constantly changing since the first Electric Lighting Act of 1882³. This Act introduced legislation that would protect customers and create a structured environment for the rapidly expanding business of electricity generation, transmission and distribution. Since then the electricity industry has gone through constant evolution due in the main to new technologies and the changing ethos on the optimal structure for the industry. This section gives an overview of the nationalised years, the motives for, and the process leading up to privatisation.

2.3.1 The early years

The Electricity Commission established in 1919 to govern electricity supply in the UK was the first of many governing bodies dedicated to the regulation and control of electricity generation and supply. It was replaced in 1926 by the Central Electricity Board which was in charge of promoting and operating a national system of interconnections between generators.

2.3.2 Pre-Privatisation

By 1947, there were 560 separate electricity suppliers of which two thirds belonged to local public bodies. This privately owned publicly controlled system of generation was difficult to regulate and lacked the co-operation needed for expansion. The Electricity Act of 1947 reorganised the whole ESI by nationalising the remaining private companies and creating 14 powerful Area Boards to control distribution, and a Central Electricity Authority (CEA) to organise finance and policy.

The ESI reorganised again in 1954 with the separation of the two Scottish Boards. Three years later three Boards were founded: the Central Electricity Generating Board (CEGB) serving England and Wales; the South of Scotland Electricity Board (SSEB) serving the region south of a line from the Firth of Tay to the Firth of Clyde; and the North of Scotland Hydro-Electric Board (NSHEB) serving the rest of Scotland including the Islands. These were made responsible for generation, transmission and distribution. The CEGB also received a small amount of its electrical energy through the High Voltage Direct Current (HVDC) interconnector with France. Figure 2.3 shows the pre-privatisation structure of the ESI.

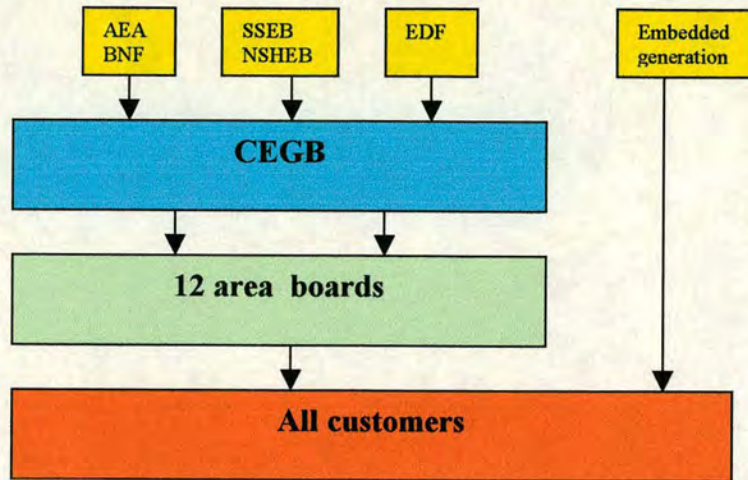


Figure 2.3. The structure of the Pre-privatisation ESI in England and Wales

The Electricity Council (EC) was set up to regulate the industry's constituent companies and provided a link to the government to help set energy policy. Generation and all other infrastructural investments were centrally planned and co-ordinated. In the earlier years, the predominant generation fuel was coal, because it was the cheapest and most readily available primary fuel source. To support the coal industry, the government prohibited the imports of foreign coal at a time when US prices were roughly half that of British coal. With 82% of its capacity in the early 1960s dependent on coal, the ESI was essentially a hostage to the coal industry. To loosen its grip, energy planning policy changed to favour other fuels. This change was also helped by low oil prices in the 1970s. By this time coal-fired capacity in the CEGB had fallen to 66% with oil accounting for 22%, nuclear 10%, gas 1.5% and hydro 0.5%⁴.

The CEGB used a merit order dispatch system, which was effectively a cost-plus pricing mechanism. The cheapest generators were dispatched first, to meet base load, followed by the increasingly more expensive units until the load was met. The CEGB charged the Area Boards a fixed amount, about £16/MWh in 1988/89, for each unit of electricity they took during the three peak half-hours of the year. The Area Boards were also charged a further £20/MWh for their average demand during the next 250 highest half-hours, to cover the higher fixed operating cost of the marginal peaking stations.

Despite the Energy Acts of 1983 and 1987, aimed at attracting private generators back into the ESI, the industry was performing badly. The CEGB was accused of bad policy making regarding the construction of nuclear plant against public opinion and poor customer service.

2.3.3 Privatisation

The mid-eighties saw a spate of privatisations, notably that of British Telecommunication and British Gas. The ideological beliefs behind restructuring and subsequent privatisation were that

- Private ownership and profit motives were better efficiency motives
- Competitive private industries gave better results than monopolies
- Government spending on long-term infrastructure costs would reduce
- Price reductions would be delivered to customers

By the 1987 elections the government's attention focussed on the electricity industry. Mrs Thatcher asked the Minister responsible, Cecil Parkinson to plan the privatisation of the electricity industry. BT and BG privatisation had been a financial success for the government but the former's quality of service had declined significantly and the latter was on the verge of being referred to the Monopolies and Mergers Commission (MMC). It was decided that, to avoid the pitfalls of the previous privatisations, structural changes would have to be made to the industry to promote competition.

In February 1988, the government published a White Paper entitled *Privatising Electricity*⁵, outlining its vision of a competitive electricity industry. This was to split the CEB into three segments

- Two generating companies: PowerGen and National Power
- A transmission company, initially jointly owned by the RECs but to be sold off separately as the National Grid Company in 1995
- The former England and Wales area boards became the 12 Regional Electricity Company (RECs) maintaining their pre-privatisation role of supply and distribution

In Scotland the situation was very different. Though the Scottish grid is connected to England and Wales via the interconnector, it is significantly smaller and was not deemed large enough to support a fully competitive system. Additionally there was the complex issue of the Highlands and Islands generation and transmission subsidies, which would be impossible to maintain under fully competitive conditions. As an alternative it was decided to create two vertically integrated companies, Scottish Power from the former SSEB and Scottish Hydro-Electric from NSHEB respectively, which later became Scottish and Southern Energy. Both were solely responsible for generation, transmission and supply and were subsequently privatised in 1991.

2.4 The Market structure – Pool

On 1st April 1990, known as Vesting day, the market was privatised. The Pool was the centrepiece of the new market, almost all generation had to be sold through it. The electricity price quoted was used as a reference for all long and short-term trades. The National Grid Company (NGC) was set up to run its operation together with maintaining the HV networks. The merit order system of the old regime, where the cheapest generators were dispatched first was retained, using the same dispatch software from the old regime. But instead of internal cost data, price bids were used to schedule the stations. This section details the workings of the pool as it then was and discusses in detail the changes and experiences brought about by privatisation ending with the experiences in Scotland.

The pool had the following features.

- All stations intending to generate had to submit bids 24 hours in advance of the quantity and price they wished to generate, together with plant availability and operating characteristics.
- The NGC used an algorithm to reconcile forecast demand with the data received from the generators to create the ‘unconstrained schedule’
- The ‘unconstrained schedule’ was constructed on a half hourly basis and consisted of the lists of plant that had to be operational to meet the demand and provide spinning reserve during the period. The scheduled plant was chosen in price order, stacking up from the cheapest first, until the expected demand for the period was met. This was used to derive the operational schedule.
- The System Marginal Price (SMP) was the accepted bid of the most expensive generator scheduled to run in a given half-hour and was paid to all scheduled plant.
- The generating capacity of the Pool for each half-hour was also calculated. This additional payment was devised as an incentive to make capacity available. Capacity payments were calculated as:- $\text{LOLP} * (\text{VLL} - \text{SMP})$: LOLP was the Loss of Load Probability, a number between 0 and 1, calculated as the probability of a power failure. VLL was the Value of Lost Load, a constant with price indexation. All available power stations would receive capacity payments whether or not they generated. The long-term value of capacity should have been relatively stable as it was designed to represent the incremental cost of adding to the system, but in the short term its value was sometimes extremely volatile.

All scheduled generators received Pool Purchase Price (PPP) calculated as SMP plus capacity payment and this was paid regardless of the level of the original offer. The cost of covering transmission losses, ancillary services and providing reserve generation was known as Uplift. PPP plus Uplift gave the Pool Selling Price (PSP), the price Suppliers paid. Uplift was not volatile and generally added around 8% to PPP⁶. Pool prices were calculated every day of the year and provisional prices were published in the Financial Times. Settlement was continuous and payments were made by NGC 28 days after the scheduled date. The structure of the post-privatised structure is shown in figure 2.4.

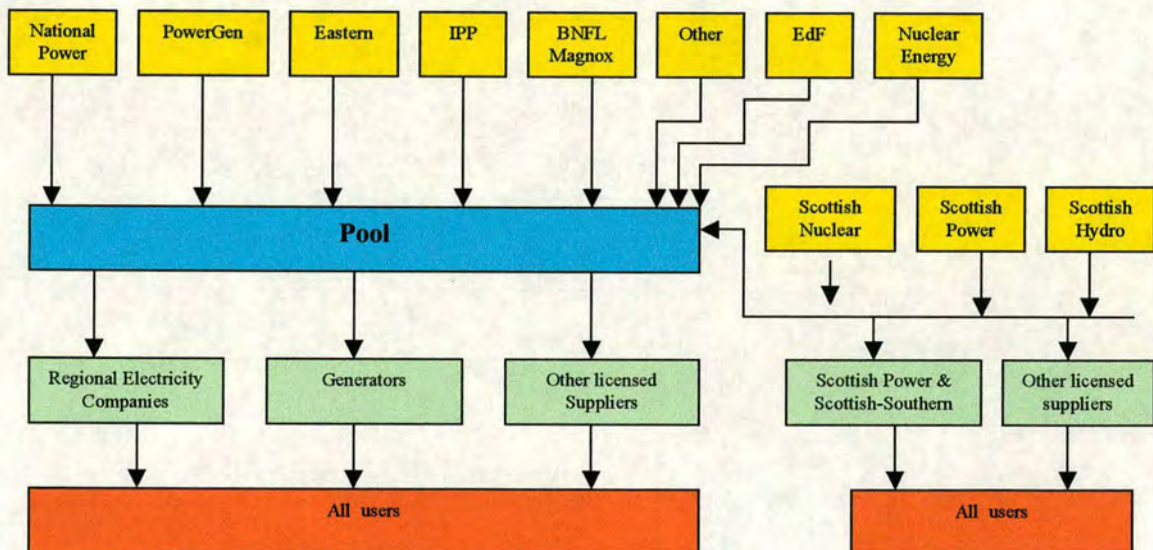


Figure 2.4. The post-privatised structure of the ESI

2.4.1 Competition in Generation

At privatisation the decision to create two generators was driven by the desire to privatise the CEGB's nuclear power stations. National Power with 30 GW of conventional plant was to be able to absorb the risk of 8 GW of nuclear plant. The remaining 20 GW of conventional plant, belonging to PowerGen was to act as a counter-balance to the larger company⁷.

The intention was to allow the two generating companies to compete for sale of their output to the RECs and larger customers. They would also have to face entry from new stations built by independent power producers.

Nuclear was withdrawn from the initial privatisation for the following reasons:

- The city was uneasy about electricity prices

- The decommissioning costs were causing concern in the city
- The memories of Chernobyl were still fresh and again the city was uneasy about litigation if a similar event were to occur in the UK

Nuclear Electric and Scottish Nuclear were sold in 1996 as British Nuclear. This was after the debt was restructured and the older stations removed from the package and retired.

2.4.1.1 The experience so far in England and Wales

This section assesses the effect of privatisation on the UK's generation profile since privatisation by detailing the experiences of the various generating fuel sources over the last ten years.

2.4.1.1.1 Gas

There has been a dramatic switch from coal-fired to gas generation. There was no gas-fired generation in 1990, 16.6 GW had been commissioned by 1998. The reasons for these changes are outlined below.

- Combined with low capital and running costs the CCGT technologies appearing by the late 1980s required much shorter construction times (24-36months).
- The expiration of the long-term 'take-or-pay' gas contracts with British Gas and the introduction of a competitive gas market led to a reduction in the price of gas fuel and has made it more instantly accessible.
- The European commission rescinded its prohibition on gas-fired generation and tightened emission limits on sulphur dioxide. To meet these limits National Power and PowerGen would have had to refit some of their coal stations with flue-gas-desulphurisation equipment. Given the relative prices of gas and coal, they decided to replace some of their coal plant with CCGT stations.
- Eleven of the RECs went into partnership with Independent Power Producers (IPP) to invest in CCGT stations. The stations sold long-term contracts (typically 15 years) for their expected output on terms that allowed them to be largely debt financed. The RECs involvement with IPPs was welcomed by the Regulator because it encouraged more competition in generation.

2.4.1.1.2 Coal

The ascension of gas was at the expense of the demise of coal. It fell from meeting 66% of the UK's annual demand in 1990 to 28% by 1999. Though this was inevitable due to the emergence

of CCGT technologies, it was partly precipitated through government policy. The main thrust of which are outlined below:

- The generators were privatised with three-year 'must-take' contracts to buy fixed quantities of British coal, at above world prices, and to sell electricity to the RECs. The prices in the electricity contracts were high enough to cover the cost of the coal and this protected the coal industry from competition.
- When these contracts expired, the new replacement contracts allowed the generators more flexibility to shop around. The reduced obligation on generators to buy over priced coal saw a shift away from coal to gas.

2.4.1.1.3 Oil

Generation by oil powered plant has also reduced from 7% to 2% of the UK's total annual demand since privatisation. This deficit has been made up by the increase in gas generation. The primary reason for this demise is its high emission levels. The most efficient Combined Cycle Oil plant would have NO_x emissions of 0.19g/kWh, which are three times higher than that from CCGT stations and SO_x of 0.25g/kWh, which is less than coal but high compared to the insignificant levels produced by gas combustion. Although the oil emissions are lower than coal, the price of coal was only 58% of that of oil on a heat equivalent basis, in 1997. This made coal fired generation cheaper than oil powered generation that year.

2.4.1.1.4 Nuclear

Since 1990 its percentage of the national capacity has increased by 35% whilst the costs have dropped by 30%. This is because

- The 'must take' status afforded electricity generated by nuclear as part of the Non Fossil Fuel Agreement (NFFO) guaranteed the spot market price for all its output.
- The bid to privatise the nuclear stations saw the closure of the less efficient Magnox stations and the writing off of the investment cost to make the industry more attractive to the private investor. The subsequent low capital cost is reflected in minimal marginal cost.

The remaining stations were sold as British Energy in 1995.

2.4.1.1.5 Others

Imports from France and Scotland have gone up since privatisation, this is due to the receipt of market price for their output. The financial incentive to increase output has resulted in the Scottish interconnector increasing its exports from 850MW at privatisation, to 1600MW

presently. This is set to be increased further to 2200MW by the end of 2001. The majority of the flows are North to South

2.4.1.2 The experience so far in Scotland

Scotland has maintained much of its pre-privatisation structure, which has made new entry very difficult. British Energy is the largest generator and provided 60% of Scotland's requirements in 2000. The remainder of the total output was roughly met by a 50:50 split between Scottish Power (SP) and Scottish Hydro-Electricity's (SSE) conventional stations.

In Scotland there was no Pool, so plants were despatched on the basis of marginal cost. But in practice the number of long-term contracts between the companies complicated the picture somewhat. These contracts are outlined below.

- **The Nuclear Energy Agreement:** This was a 'take-or-pay' contract between British Energy and the two generators, SP and SHE. It provides 60% of Scotland's requirements and prevents SN from selling any of its output to other parties without consent from the two generators
- **The coal agreement:** This gave SHE access to one-sixth of Scottish Power's coal fired plant
- **The Peterhead agreement:** This gave Scottish Power title to acquire 50% of the present capacity from SHE's Peterhead station
- **The Hydro Agreement:** It was a 'must-take' that gave Scottish Power title to 400GWh per year of SHE's Hydro-Electricity output, subject to rainfall conditions.

These contracts, coupled with Scottish Power and SHE's control of the interconnector access and pricing greatly limited the potential for new entry. As a result, the structure of the Scottish generation industry has not changed much since privatisation⁸.

2.4.2 Competition in Transmission

At privatisation it was decided to create one entity, National Grid Company (NGC), to be responsible for system operation. The following measures were imposed to limit the abuse of its monopoly position.

- Transmission price controls were imposed on the NGC to limit its monopoly power.

- No UK ESI company was allowed to hold more than 1% of the shares in the NGC group.

The National Grid Company retained all its pre-privatisation responsibilities. In Scotland, Scottish Power and SHE controlled and operated the transmission network in their respective areas.

2.4.3 Competition in Supply

Competition in Supply was deemed essential to a fully competitive electricity market; shown by its high emphasis within the 1989 Electricity Act. Central to this was the gradual opening up of the Suppliers captive customer base to competition. The theory was that consumer choice in Supply would put pressure on them to buy electricity more cheaply, which would put downward price pressure on generators.

At the time of privatisation, Suppliers were split into the following two categories:

- **First-tier Supplier:** These were the RECs which were created at the time of privatisation, from the 12 area boards in England and Wales, the two Scottish companies and Northern Ireland Electricity. These companies owned the distribution business within their regional area and were obliged under their licence conditions to supply all customers in their region with electricity. The REC's dual functions of distribution and supply had separate accounts and were regulated differently. With 22 million customers between them, the RECs were the biggest purchasers of electricity on the wholesale market. The total value of their franchise markets was about £7bn.
- **Second-tier Supplier:** These were defined as the companies that supply electricity to customers within another RECs regional area. The players in this market ranged from RECs, Generators, and independent Suppliers to large customers. Second-tier suppliers had to pay the first-tier supplier a fee for use of the distributional network.

Both first and second tier Suppliers had to sign a number of statutory codes and agreements, including the Grid Code (GC), the Distribution Code (DC) and the Pooling and Settlement Agreement (P&SA)

Full competition was to be achieved over an eight-year period with successive customer tranches given the option to shop around for their electricity. This was phased in three stages, known as the *franchise break*, outlined below.

- **Stage 1:** This took place at the time of privatisation. All customers with peak demands above 1 MW were allowed to choose their Supplier from any licensed Suppliers within the UK. The number of customers in this group was about 5000.
- **Stage 2:** The second phase occurred on 1st April 1994. This group comprising around 50 000 were customers with peak demands between 1MW and 100kW.
- **Stage 3:** This was originally scheduled to happen on 1st April 1998, but the process only began in September 1998 and was completed on 24th May 1999. This opened up the domestic sector involving 26 million customers across the UK to full competition.

Suppliers are free to buy power from whosoever they choose. Some even went into partnership with IPPs to meet some of their customer demand.

2.4.3.1 The experience so far in England and Wales

The RECs lost two-fifths of their sales volumes in the first year of privatisation. Their market share in the above 1MW segment continued to decline from 57% of the market in 1990/1 to 20% in 1998/9 shown in Figure 2.5a. Similarly, there was a noticeable loss of 10% of market share in the 100kW to 1MW market between the financial years 1997/8 through to 1998/9⁹.

The final phase of supply deregulation was completed on 24th May 1999. By March 2000, slightly under a year after full deregulation, just over 4 million electricity consumers (16½%) were no longer with their local REC. In the domestic market segment, new Suppliers were charging on average £10 a year less than the incumbent Supplier. In some areas, annual price reductions offered by a new Supplier can be in excess of £60. In addition to competition, price controls imposed by the Regulator has reduced electricity prices for domestic and small business customers in real terms by about 7% on average from 1 April 1998 and a further 4.4% from April 1999. These reductions were expected to save a typical domestic customer around £45 over the period from April 1998 to April 2000. A typical annual UK electricity bill was £264 in 1999 compared to £299 in 1995¹⁰.

However the number of complaints about poor sales techniques, Supplier poaching and inaccurate billing has increased.

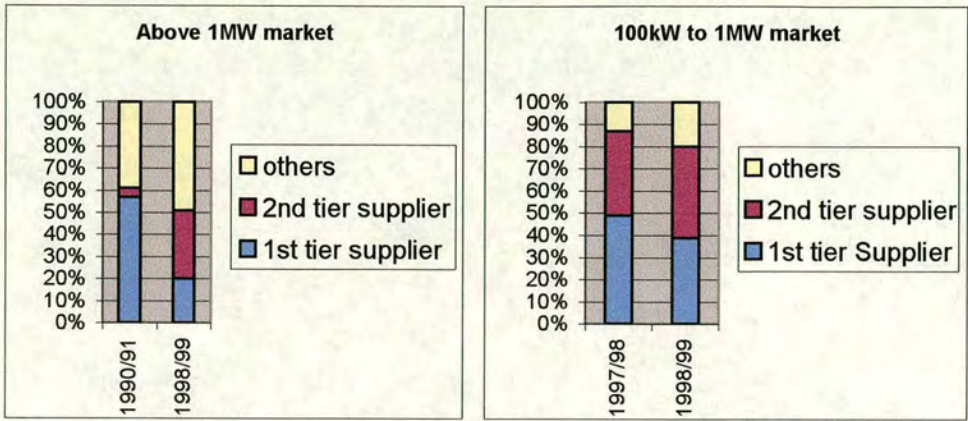


Figure 2.5a and 2.5b. Decline in 1st Tier supplier market share

2.5 The Present situation

Against all expectations the price of electricity to the customer has gone up since privatisation. Prices should have fallen for the reasons laid out below.

- Competition was intended to make the industry more efficient, making prices more reflective of the marginal cost of generation.
- In theory the lower cost of generation brought about by reductions in fuel costs, lower capital cost of new plant, the introduction of newer, cleaner and more efficient technologies should have been reflected in consumer prices.

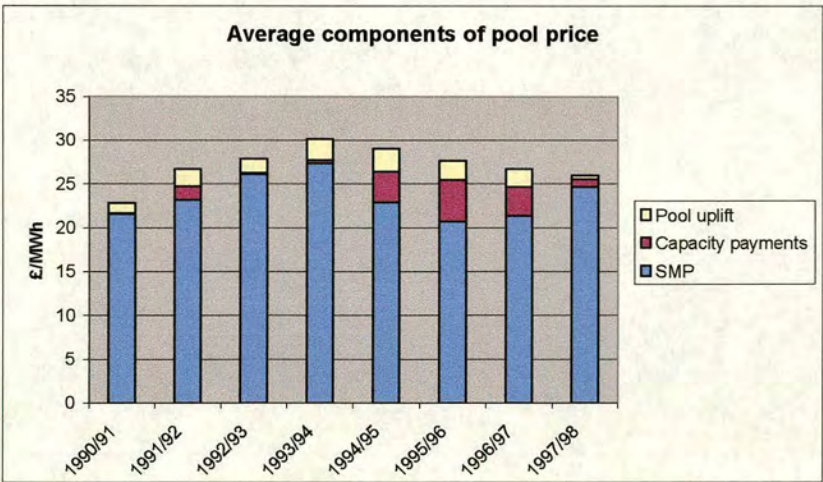


Figure 2.6. Annual average time-weighted pool prices (February 1998 £/MWh)

To investigate the reality, figure 2.6 shows the average annual time-weighted Pool prices since privatisation. From this graph, prices had indeed increased, the temporary lower prices between 31st March 1994 and 31st March 1996 were due to the price cap of £25/MWh imposed by the regulator on System Marginal Price (SMP). Since the removal of the cap the SMP steadily crept up¹¹.

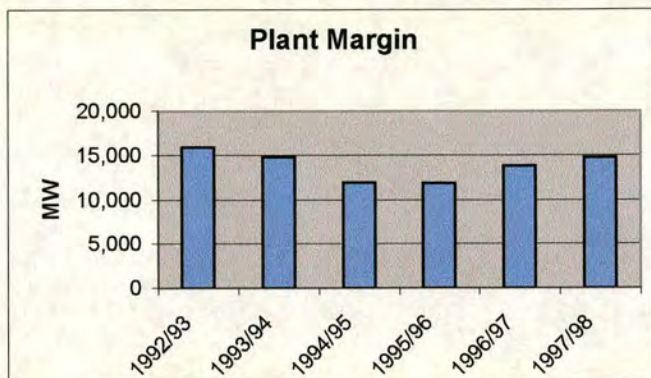


Figure 2.7. Average winter plant margin (October to March)

It was possible that the increased SMP was reflective of the lack of available plant and or increased demand. This was investigated in figure 2.7. The graph shows that the average winter plant margin, the difference between total plant availability and demand fell up to 1994/5 and then increased again. This suggested that the increases in SMP had no relation to increased demand or reduced capacity.

The evidence from the preceding discussions suggested that the market was not behaving as intended. The issues of market structure and market power will be examined in the following sections to determine the source of this distortion. For the purpose of the discussion, the power market must be thought of as three integrated and interacting markets. These markets model the structure of electricity trading in the UK. There are

- A physical market for spot energy (the Pool)
- Markets for risk-sharing (the contracts and Electricity Forward Agreement markets)
- A market for capacity.

2.5.1 Power market games and strategy through the Pool

In order to identify the reasons for the increases in SMP, the price-setting mechanism was considered. SMP was set by the mid-merit and peak load stations because they were the upper limit of accepted bids. The stations that came into this category were the coal, oil and the more expensive gas sets.

However, it was immaterial who set SMP, since all the Generators received it, and thus all benefited from any increases. Whether the increases in bid prices yielded the desired increase in revenues depended on the confidence of the price setting generators PowerGen and National Power, that not too much additional capacity would be brought into play by their competitors. It is therefore necessary to look at Generators changes in output and capacity since privatisation.

Figure 2.8 shows the Generators percentage of output in England and Wales. From the figure it is obvious that all the Generators increased their output except the largest two, National Power and PowerGen who reduced their output by retiring about 17000MW of old capacity and disposing of 6000MW of coal-fired plant to Eastern Electric. The latter was replaced with 6000MW of new CCGT capacity. As a result their combined market share fell from around 78% to 41% of output and 51% of capacity.

In a competitive market increased bid prices not reflective of marginal cost will result in competitors increasing their output. Price increases could only be sustained if those wishing to secure them were willing to lose a share of the market. Similarly, Companies wishing to increase their market share will have had to reduce their bid prices, unless some other participants were willing to trade a loss of market share in order to sustain prices¹².

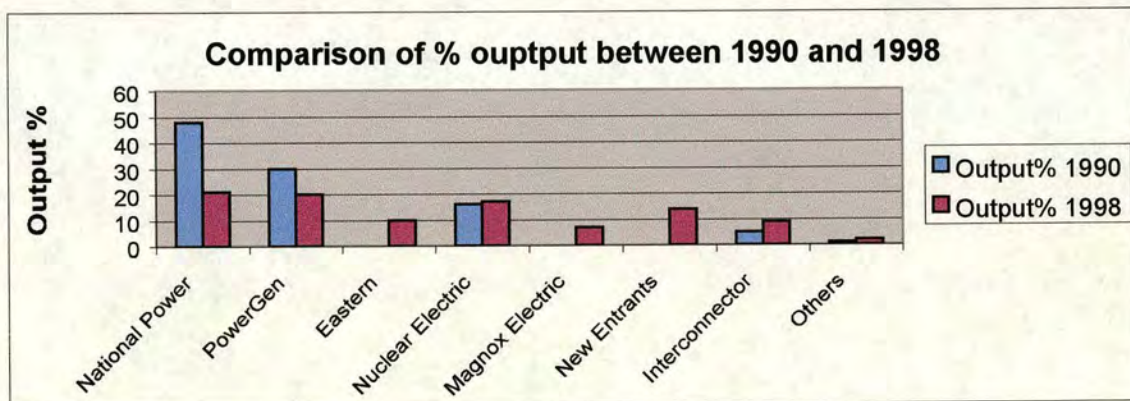


Figure 2.8. Market shares of output in England and Wales

The above discussion suggested that National Power and PowerGen were content to lose a percentage of their output to facilitate the increase in SMP. The other Generators seem content to benefit from the increased share of output, and higher SMP.

	Output (TWh) winter 1996/97	Output (TWh) winter 1997/98	Change (TWh)	Change (%)	Average winter load factor
National					
Power	26.6	23.2	-3.4	-13	45
coal	7.4	10.3	+2.9	+2.9	93
CCGT	0.4	0.1	-0.3	-0.3	2
Oil					
Total	34.9	33.6	-0.9	-3	46
PowerGen					
Coal	19.7	18.8	-0.9	-5	46
CCGT	1.9	12.2	+0.3	+3	90
Oil	1.4	0.0	-1.4	-98	0
Total	33.1	31.1	-2.0	-6	47
Eastern					
Coal	12.2	15.6	+3.5	+29	61
CCGT	1.5	2.5	+1.0	+68	73
total	13.7	18.2	+4.5	+33	63

Table 2.1 winter output and load factors by plant type

The next issue of concern was discrimination according to fuel type. Table 2.1 shows the winter output and load factors by plant type of the largest three generators. From the table, it may be seen that the two SMP setting generators reduced the output of their coal stations, whilst increasing their CCGT output. These figures suggest that the big two Generators were following a strategy of profitable withdrawal from coal-fired generation. The increased output of the CCGT stations went in at baseload, thereby receiving the SMP set by their more expensive stations. The sum-total effect was increased revenues from the higher SMP received by the CCGT plants and reduced fuel and environmental costs.

In contrast, Eastern increased their coal generation output but it seemed that they were not actively competing with the big two in setting SMP but were willing to bid lower to get the increased output and receive SMP.

The discussion in this section endorsed the suggestion that the big two were using their market power to take advantage of the market structure to realise increased revenues and that the marginal pricing mechanism did not encourage the other participants to compete on price. The smaller participants benefited from increased prices and increased market share, with consequent increased revenues.

2.5.2 Power games and strategy through the contracts market

In practice, between 80 and 90% of electricity sales were hedged with Contracts for Difference (CfDs)¹³. With a two-way CfD, the parties agreed a strike price for a fixed quantity of electricity. Whenever the pool price was below the strike price, the buyer paid the seller the difference

between the two. Conversely when the pool price was higher, the seller refunded the difference. Theoretically, Generators with high CfD cover, were indifferent to the Pool price because their income became the strike price. However in practice things were somewhat different with the Pool price being used to manipulate contracts. The most commonly used strategies are described below

- If the Pool price was below the station's marginal cost, the Generator would buy electricity from the Pool to meet its demand commitments, and not trying to run
- In the medium term, contract prices depended on expected Pool prices, so increasing this strengthened the Generators bargaining power in contract renegotiations.
- Increasing price volatility (attributable to increased capacity payment volatility) made contracts more attractive because they guaranteed predictable electricity costs. This increased the demand for CfDs thereby pushing up their value.

2.5.3 Power games and strategy through the capacity payment mechanism

Capacity payments were designed as an incentive mechanism to reward Generators for providing capacity. At times of low demand and excess capacity, the capacity element was zero. These payments increased as demand rose and generating capacity got more stretched. The long term value of capacity should be relatively stable since it represented the incremental cost of adding capacity to the system, but the short-term value can be very volatile. Proponents of the capacity mechanism system argued that this system provided valuable price signals to the market. These signals acted as the incentive to build more capacity as and when the market needed it.

In reality the reverse was the case as is shown by the inverse relationship between capacity payments and SMP in figure 2.6. In the years when SMP was low, capacity payments were high, the inverse happened when SMP was high. For example, during the period of the SMP cap, capacity payments rose as a result of the largest two generators withdrawing capacity from the system. The generators managed this by increasing their plant failure rate. This reduced the level of spare capacity on the system, raising capacity payments. This inverse relationship suggested that the capacity payments were not working as intended, but instead were being used as a means for topping up Generator income when SMP was low^{14 15}.

2.5.4 Discussion

The issues of markets power were focused on in the earlier sections. It was felt by many that the Pool was not working as intended because the large Generators appeared to be using their market power to take advantage of the market structure to make excessive profits. This prompted the

Government to instruct the Regulator to investigate the possibility of a new trading structure in the UK.

2.6 Summary

The Electricity Supply Industry (ESI) is responsible for the production of electricity from primary fuels and transporting it to the end-user. The industry has undergone many structural changes over almost a century and a half of its existence culminating in re-privatisation in 1990. This was intended to bring cheaper prices to customers and lead to a more efficient utilisation of the generation resource through competition, but these were not achieved. Instead, prices have risen in real terms, and Generators have appeared to abuse their market power and the pricing mechanism. This prompted the Labour party, shortly after its re-election in May 1997, to instruct the Regulator to investigate the possibility of new trading arrangements for the industry.

3. The New Trading Arrangements (NETA)

3.1 Overview

In this chapter, the New Electricity Trading Arrangements (NETA) are described in detail, followed by a summary of the British Electricity Trading and Transmission Arrangements (BETTA), the proposals for the extension of NETA to include Scotland. The most probable effects on the Electricity Supply Industry (ESI) are discussed, together with an analysis of the specific impact on the various groups of participants. The chapter concludes with a focus on Suppliers, the group with the lowest margins within the ESI. They are presented by NETA with the steepest learning curve, making them the industrial segment most vulnerable to the changes. As a result, Suppliers will have to be able to measure and manage their exposure to risk if they are to survive in the new era of trading.

3.2 NETA

In October 1997, shortly after Labour's election victory, The Minister for Science, Energy and Industry, John Battle, instructed the Electricity Regulator, Professor S C Littlechild to investigate reforming of the electricity trading arrangements. The initial proposals were published in the *Review of Electricity Trading Arrangements: Proposals* in July 1998¹⁶. The proposed trading arrangements were modelled on the gas market and competitive electricity markets in Scandinavia and Australia. These were accepted by the Government in October 1998 and included in its White Paper on Energy Policy. It was confirmed that the Office of Electricity Regulation (OFFER) and the Department of Trade and Industry (DTI) would lead the process of consultation and with the help of a Programme Director, facilitate the full participation of the industry and its customers. After much consultation *The New Electricity Trading Arrangements* (NETA)¹⁷ were published in July 1999, and introduced on 31st March 2001.

The aims of the proposed trading arrangements, laid out by OFFER were as follows:

- Lower prices from more efficient and competitive trading
- Greater choice of markets
- More scope for demand side management
- Sharper incentives to manage risk
- Transparency from simpler bids
- Forward price curves to facilitate new entry

Chapter 3. The New Trading Arrangements (NETA)

- Avoidance of discrimination against fuel sources by rewarding flexibility
- More liquid contracts market
- Scope for greater co-ordination and consistency with gas
- More flexible and effective governance

The system of marginal pricing, together with market power were identified as the major obstacles to competition. The then marginal pricing system was to be replaced by a series of bilateral markets operating from the long-term up until real-time. It was perceived that bilateral markets, based on the gas markets, would be less open to the abuse of market power and less vulnerable to gaming. The main thrusts of these markets at the time of writing are outlined below:

- Forwards and futures markets (to the extent required by participants)
- Short-term power exchanges (also to the extent required by participants)
- A voluntary Balancing Mechanism (initially operating from 4 hours ahead to the end of each trading period)
- A mandatory settlement process.

It should be noted that the above aims might be liable to some degree of change over time. It was expected that trading would commence in the long-term via either exchange-traded derivatives or by long-term bilateral contracts known as Over-The-Counter (OTC) contracts. These were expected to allow participants to cover their expected long-term positions. These markets will operate from any time in the future up until 24 hours before the half-hour in question. They will serve as indicators to the short-term markets and in reverse, the short-term markets will provide the indices on which the long-term markets will be referenced. The short-term bilateral market will be functional from 24 hours till 4 hours before the period. The balancing market will be in operation from 4 hours until real-time. The settlement process takes place after the scheduled period. Figure 3.1 is a schematic diagram of how the three markets interact with each other.

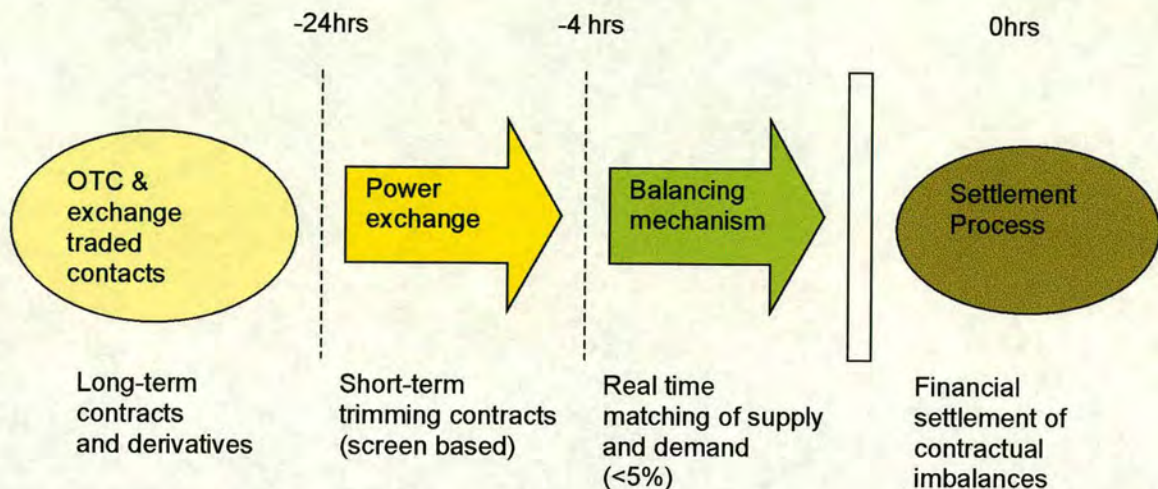


Figure 3.2. The time-periods for NETA's physical markets

Each of these markets is discussed in more detail in the following sections.

3.2.1 Long-term markets

It is expected that the majority of trading will take place in the forwards, futures and long-term bilateral markets. These are expected to evolve according to market participants needs through third party involvements. Trading via long-term bilateral contracts or forwards will allow participants to cover their likely output or demand so that by the time the Balancing Mechanism opens for a trading period, participants contracted positions will closely match their anticipated physical positions. The level of cover required by participants will depend on their trading strategies and hedging requirements

3.2.1.1 Forwards

These are expected to trade like financial forwards. The buyer and seller agree on a price one day, on a quantity of electricity for future delivery. On the delivery date, the buyer pays the seller, or the seller pays the buyer the difference between the agreed forward price and the current market price also known as the spot price. Through this, market participants will be able to secure their long-term output or consumption, well in advance. In this market, the terms and agreements of all contracts will be determined bilaterally. In other words it is an Over-The-Counter (OTC) market.

3.2.1.2 Futures

Futures are similar to forwards in that one party agrees to buy electricity today for the future physical delivery from a second party. However, there are the following differences between futures and forwards

- Futures are standardised contracts and are traded on an exchange. The International Petroleum Exchange (IPE) has already expressed an interest in launching an electricity futures market¹⁸.
- Participants in the futures market are not subject to counterparty default risk because the exchange acts as counterparty, guaranteeing the settlement of all transactions.
- Trades are settled at the end of each trading day. This means, instead of settling the trade at the end of the contract as with forwards, the difference between the future price and the spot price is calculated and settled at the end of each trading day, this is known as the margin. The margins are accumulated and settled on the final agreed date of physical delivery.

It is hoped that price reporting will evolve within these markets. This would give an indication of the market's perception of 'future' spot prices, also known as price transparency. Experience from the Australian and Scandinavian electricity markets suggest that this will lead to more efficient competitive pricing. Greater transparency will increase third party participation in these markets whilst increasing the frequency of trades. The increased volume of trades is often referred to as liquidity.

3.2.2 Short-term bilateral market

The short-term bilateral market will operate from about 24 hours up until 4 hours before the trading period. It will be operated via a screen-based system, price setting will be auction based as opposed to marginally priced. The easy access to bidding information via screens, should increase transparency and maintain anonymity. As at the time of writing the proposed market operator was UK Power Exchange (UKPX). It would act as underwriter for all contracts in return for transaction charges.

Generators, suppliers, customers and traders will submit simple offers and bids to the market operator. Offers and bids will be posted, modified or withdrawn at any point until they are accepted. Accepted offers and bids will represent firm financial commitments and will be settled at the prices specified in the offer or bid, rather than some uniform price specified by the market operator.

Trading will be on the basis of standardised products. For example, baseload contracts, peaking contracts and combinations of these contracts as well as individual trading period contracts. The nature of these products will be allowed to evolve over time to meet the needs of market participants.

This short-term bilateral market will thus be similar to exchange-based markets for shares and commodities. It is possible that this market will not only come under the regulation of licensees by the Developmental General of Electricity Supply (DGES), but would also be subject to regulation by the Financial Services Act (FSA).

3.2.3 Balancing Arrangements

The Balancing Mechanism is used by the System Operator (SO) to keep the system in balance on a real-time basis, both half-hourly and instantaneously. This involves resolving transmission constraints, placing Generators on stand-by to provide balancing services and ensuring that all transmission equipment will continue to operate within safe limits. NGC has estimated that the balancing activities may, on average, require around 2 GW, which is about 7% of average demand¹⁹.

The balancing market starts operating at Gate Closure (GC), which has been set at 4 hours until the period of operation. There is an obligatory requirement on all market participants to submit their intended generation and demand to the SO by GC. This information together with the NGC's own forecast are used to determine possible short-falls or excess generation or demand at various locations on the network, known as Balancing Mechanism Units (BM Units). Participation in the balancing market will be voluntary except at times of system stress. Participants in the balancing mechanism will have to comply with the Balancing and Settlement Codes (BSC) and any relevant Grid Code obligations.

3.2.3.1 Balancing points on the network

Since demand and transmission constraints are location specific, the SO will need to accept bids and offers at specific points on the transmission system. The generic term 'BM Unit' has been introduced to define the type and size of grid point location at which balancing will take place. Its definition is as follows.

- Connectable generating plant of capacity larger than 50MW
- Meters at the Bulk Supply Point (BSP) for customers directly connected to the grid
- Any other generation or demand connected directly to the transmission network will have its BM Unit as the meter at the BSP below which the meter is connected.
- For smaller generation or demand that might be connected to the distribution network rather than the transmission network, the BM Unit will be an aggregation of all meters within a BSP group for which the participant is responsible (including those for any licence exempt Generators with which the participant has signed a contract)
- For interconnectors the BM Unit will be the proportion of the interconnector meters allocated to each Interconnector user.

There are some categories of demand that will need special treatment, for example radio-teswitched customers. This will be discussed in more detail in subsequent chapters.

3.2.3.2 Balancing Mechanism (BM) operation

Participants must submit their Initial Physical Notification (IPN) to the SO by 11:00am for the day ahead. The IPN shows the Generators or Supplier's expected position for each half hour period in the next day. The IPN's are constantly updated until Gate Closure when participants must submit a Final Physical Notification (FPN), firm declaration of output or demand during the specified half-hour.

At gate closure, the balancing mechanism opens. A wide range of participants are able to make Bids and Offers to the SO through the Balancing Mechanism. Bids and Offers are defined below.

- **Offers or 'increments':** Increase in output from Generators and decreases in consumption from Suppliers
- **Bids or 'decrements':** Load-reductions by Generators and consumption increases by Suppliers

It was proposed that 'buyers' of imbalance electricity would pay a price calculated as the volume-weighted average of the Offers accepted in the Balancing Mechanism known as System Buy Price (SBP). Sellers of imbalance electricity will be paid the volume-weighted average of accepted Balancing Mechanism bids, referred to as System Sell Price (SSP).

In general, Generators who are under contracted (and Suppliers who are over contracted) and 'spill' electricity on to the system, can expect to receive the lower SSP for their electricity. This is expected to encourage participants to resolve their imbalance in the forward markets. Suppliers who remain under contracted (and generators who under generate) can similarly expect to be charged the higher SBP which is expected to be higher than if they had entered into contracts for their full requirements in advance. These charges are expected to reflect the additional costs incurred by the SO in instructing participants at short notice to keep the system in balance from moment to moment. The costs of any forward contracts used by the SO to maintain a balance of overall supply and demand are also included in the calculation of imbalance prices.

Bids and Offers are location specific and refer to individual BM units to help the SO overcome local transmission constraints.

3.2.4 Cash-Out and Settlement Arrangements

Imbalance cash-out prices are designed to penalise or compensate Generators and Suppliers on whose behalf balancing services are performed by the SO. Electricity imbalances are the difference between participants metered volumes and their contracted volumes determined from their FPNs and balancing market activities. These are charged SSP for over contracting and SBP for under contracting. The settlement of imbalances is to ensure that any electricity not covered by contracts or BM actions, is paid for or charged an appropriate price.

Default imbalance prices will be calculated if there are no Offers or Bids in a relevant half-hour period. Options for this are outlined below:

- Price from the previous half hour
- An average of all cash-out prices over the past 7 days
- The price of the offer/bid that would have been taken first (the lowest priced offer or the highest bid)
- A price indicator from the spot market , adjusted by a transaction fee

3.2.4.1 Imbalance Volumes

The volume to which imbalance prices are applied is the difference between a participant's notified contract volume and metered volume. A participant's notified volume is the sum of all trades (whether over-the-counter or exchange-based) notified the SO by the time of Settlement, together with any volumes accepted by the SO in the Balancing Mechanism.

Imbalances relating to production (export) and consumption (import) meters will be calculated separately. All the contributions from production and contribution meters associated with a participant will be aggregated, regardless of their geographical location.

3.2.5 Regulation

As mentioned earlier, OFFER was set up in 1989 to oversee the regulation of the ESI, and its control over the electricity industry was through licence agreements. However, the Regulator could not change the licences without the holder's consent, so new regulation was slow to achieve. Rapid intervention was only possible through the Monopolies and Mergers Commission or directly through legislation from the Secretary of State. This resulted in a call for more efficient regulation, consistent with the overhaul of the industry.

On 19th June 1999 the merger of the two energy regulators, OFFER and OFGAS, became the Office of Gas and Electricity Markets (OFGEM). Because of the converging interdependency of the two industries, this was intended to improve the consistency of gas and electricity regulation.

OFGEM has increased powers under NETA. This is achieved through shorter licensing periods and some direct control over the market controller so as to enable quicker regulation through market changes. It is hoped that regulation through the market rather than by direct intervention will encourage more self-regulation via competition.

3.2.5.1 Governance

The futures and short-term bilateral markets are likely to be subject to regulation under the Financial Services Act. The Balancing and Settlement Code (BSC) for the balancing markets and the Imbalance Settlement Process (ISP) will replace the Pooling Settlement Agreement and SO will be obliged to maintain and enforce the Code. There is a Balancing and Settlement Code Panel, comprising representatives of all interest groups including Generators, Suppliers, traders and customers as well as the SO. This panel considers modifications to the rules and procedures, and listens to representations from any concerned parties. All proposed modifications to the BSC will be subject to the approval from DGEN.

3.3 Scottish Trading Arrangements: British Electricity Trading and Transmission Arrangements (BETTA)

The Scottish market is an important part of Great Britain's electricity market, both in terms of volume and in diversity of generation sources. Annual electricity sales in Scotland are approximately £1bn and form an important component of the national turnover²⁰.

The structure of the Scottish market has evolved very slowly since privatisation. This was thought by some to be due to the continued vertical integration of the two principal companies, the restructuring of contracts, and interconnector access and pricing. The Review of Scottish Trading Arrangements (ROSTA) published in October 1999 began the process of consultations designed with the intention of establishing new Scottish Trading Arrangements, separate, but consistent with NETA. This document outlined proposals for interim trading arrangements to be put in place between the expiry of the previous arrangements in April 2000 and the introduction of the new arrangements in April 2002²¹. The May document, *Initial proposals and issues for the consideration on the reform of Scottish Trading Arrangements May 2000*, included more detail of the vision.

In August 2000, the *Interim Proposals for the reform of Scottish Trading Arrangements: British Electricity Trading and transmission Arrangements (BETTA)* was published. In this document, the initial plan for separate Scottish Trading Arrangements (STA), set out in the May document, was abandoned in favour of nationwide trading arrangements. The document proposed the following:

- Trading arrangements for Scotland as part of a Great Britain (GB) market
- Transmission access and charging arrangements for Scotland to be developed as part of a GB set of arrangements, BETTA, set to go 'live' in April 2002
- A set of interim trading arrangements to provide a smooth transition between the present set-up and the new arrangements.

The interim proposals which will run from 1st December 2000 to 1st April 2002, are designed to improve the competitive position both within Scotland and between England and Wales, before the introduction of BETTA. The interim measures are as follows:

- Encourage new supply entry by modifying the existing Nuclear Energy Agreement (NEA) to allow British Energy to sell some of its capacity to other Suppliers or directly to customers.
- Increase the transparency of the interconnector access arrangements, and define the cost of interconnector trading.
- Balancing services will be charged at the proposed balancing use-of-system charge in England and Wales, less the constraint related component.
- Imbalance prices in England and Wales will be used as a proxy for imbalance prices in Scotland.
- Initial proposals in the May 2000 document to have a single Scottish System Operator combining the existing system operation functions of SSE at Port-Na-Craig and SP at Kirkintilloch.

Since BETTA will be an extension of NETA to Scotland, all the NETA market mechanisms described in previous section will apply to Scotland from 1st April 2002. Under BETTA there will be one system operator for the whole UK. As at the time of writing, there were still several pertinent issues that have still to be resolved affecting interconnector pricing and access.

3.4 NETA's Implications on the ESI

NETA will have a marked effect on all participants within the ESI because of its change of market emphasis. The main industry wide changes include:

- Greater demand-side participation
- A more efficient competitive electricity market
- Removal of fuel source discrimination
- The ability to respond effectively to changing market circumstances through more efficient governance

3.4.1 The 'NETA effect'

The 'NETA effect' is the term used to describe the anticipated changes and effects NETA might have on the ESI

3.4.1.1 Bilateral versus Marginal pricing

One of the main reasons for the abolition of the Pool was to bring an end to the marginal pricing system. The perception was that a Generator who was paid more than he asked for was getting too much, since he was free riding on high prices set by other participants. It is still not clear whether paying stations their bid price via bilateral agreements will lead to reduced prices.

The literature on auction theory²² suggests that, in bilateral markets, those who used to submit very low bids under the marginal pricing system would be likely to submit bids closer to the highest accepted bid, that is, if they have some idea of the likely value of the highest accepted bid. This could include Generators with several bidding stations, because knowledge from all their stations could give them some foresight of the highest accepted bid. For smaller plant who were benefiting from the marginal price set by others, they would try to guess the highest accepted bid. If their guesses are wrong this could result in real market inefficiencies with high marginal cost plant being run before plant with low marginal costs. Consumers will be, in effect, paying too much if expensive plants are being run while less expensive plants lie idle.

This system could favour larger portfolio Generators with information about the bid prices of their own plant. Smaller Generators with no 'insider information' might have to sign long-term contracts to guarantee their output at less favourable prices. This phenomenon could discourage new entry from smaller Generators, and encourage larger firms to build new capacity. Eventually, this could lead to less competition as a result of fewer companies, leading to higher prices²³.

3.4.1.2 Abolition of capacity payments

Capacity payments are one of the most disliked features of the Pool and critics argue that it is an easy gaming mechanism. Regardless of the market configuration, there will always be a need for peaking stations which are used infrequently for providing capacity at the margins. Capacity payments offered them a means to recover their costs. The abolition of the capacity payment mechanism will remove the financial incentive for Generators to make extra capacity available. This would increase volatility because the system will be operating within tighter margins.

This gives very flexible plant such as CCGT greater market power in the short-term and balancing markets. For example, flexible plant could withhold capacity from the balancing mechanism until the last minute, as such raising the value of capacity and thus benefiting from a high imbalance price. This could also have a knock-on effect on the price of long-term contracts

because market participants, especially Suppliers, will want to lock into long-term contract to hedge against the spot price volatility.

3.4.1.3 Demand-side participation

Demand-side participation in the price setting process is seen as a good thing, because no longer will Suppliers have to accept, passively, the price set by Generators. They could have a choice as to what they pay. However, only Suppliers with demand-responsive customers would truly benefit from demand-side bidding. Demand-responsive consumers could react to high prices by shedding load, reducing the generation and reserve required. This could imply lower peak prices, but for a longer period. On the other hand, Suppliers who do not have price responsive customers cannot alter the amount of generation required. Generators could react to this by reallocating their sales between markets, withholding capacity in the short-term markets, raising prices, thereby negating the ethos of demand-side participation²⁴. Suppliers with high domestic loads would fall into the latter category.

3.4.1.4 Possible developments of financial markets

The Government and OFGEM expect sophisticated financial markets to evolve within the ESI to meet market participants' demands for a wider variety of risk management tools. Central to the evolution of these markets is transparency. There have been suggestions that the lack of participation by third parties in the electricity forward markets resulted in low liquidity, which meant that there was no demand for standardised products²⁵. The absence of third parties was because the Pool price, the price forward contracts are tied to, could be manipulated by the dominant Generators. As has been suggested in the previous section, if the bilateral prices are less transparent than the present system, which seems rather likely, then there would be no incentive for third party participation. The possible implication of this would be that many market participants, notably Suppliers, would face more risk under NETA and with limited tools to manage this risk.

3.4.2 Implications of NETA on the various groups

So far, the industry-wide implications of the 'NETA effect' have been outlined. The 'NETA effect' will affect the various segments of the ESI in different ways. The following section summarizes the 'NETA effect' on the various participants²⁶.

3.4.2.1 Large portfolio generators

These are the Generators with a portfolio of stations across the country, including mid-merit, peaking as well as baseload plant. Their possible operations will be as follows:

- They might be able to exercise their market power with their 'insider knowledge' in short-term bilateral.
- They will use the balancing mechanism for the margins, as they do not have much flexible plant.
- They might want cover for any imbalances.
- They are in a stronger negotiating position in long-term contract negotiation.

3.4.2.2 Nuclear

These comprise all the nuclear stations belonging to BNFL Magnox and British Energy. Their possible activities under NETA are summarised below:

- Nuclear is very inflexible and all its output will have to be fully contracted.
- Because of nuclear's inflexibility, it will have a limited role in the balancing mechanism.
- Unlike under the Pool system, nuclear Generators will not be passive price takers but will have to actively bid their output.

3.4.2.3 Independent Power Producers (IPPs)

IPPs formed the first generation of independent gas-fired CCGT plant in England and Wales. In majority of the generators, the financial loans for construction of their plant was linked to the financial return from the projected 15-year output from their stations. The power prices were linked to their price of gas under their fuel purchase agreements and also to projected Pool prices. Their position under NETA will be as follows:

- Complete contract re-negotiation will be required since their output contracts were linked to Pool price.
- They have very flexible plant and so will have the choice of under contracting so as to use the balancing mechanism to make more money for their output.
- They might become part of a portfolio, to minimise the risk of under-contracting.

3.4.2.4 CHP, Renewables, Embedded Generation

In the government's proposed Utilities Bill it was stated that 10% of generation must be met by 2010 from renewables. This is in line with its commitments to the Kyoto protocol²⁸. This will present much longer-term challenges for NETA. Because where there is a strong financial incentive for firmer offers and bids, there will be a disincentive for participants to be involved with renewables. This is because of the intermittency of the energy sources, which means that their generating plant output is less predictable than conventional ones. As a result there would be no major incentive for new inward investment, or industrial funding of research into the long-term technological advancements.

Recent renewables projects developed in England and Wales have typically signed contracts for their output with RECs under the Non Fossil Fuel Obligation (NFFO) arrangements. The NFFO Generators are paid a fixed price for their output by the REC, who in return, is compensated for the difference between the NFFO contract price and Pool Selling Price (PSP). The compensation paid is recovered from consumers via the Fossil Fuel Levy. NFFOs will be abolished under NETA, and will probably be replaced by an obligation on Suppliers through 'Green taxes'. However this was still to be finalised as at the time of writing.

The probable 'NETA effect' on renewables will be

- An alternative reference price will have to be quoted for contracting.
- They can expect to be fully contracted, as present.
- They might form part of a generation portfolio, to limit their exposure to the balancing markets
- Because of the penal imbalance charges, there will be a greater incentive to get their forecast and control right.

3.4.2.5 Suppliers

Greater demand-side participation means that Suppliers will be more vulnerable to customers whose demand is liable to variation. This inherent uncertainty may cause Suppliers to be reluctant to commit to buying specific quantities of power, leading to conservative bidding at the margin, which could push prices up. Suppliers who do not take steps to control their risks may lose business to those who understand the requirements of their customers and who introduce incentives both in their prices and their terms of trade that allow greater predictability of customer response. The effects of NETA on Suppliers are outlined below.

- Required to submit FPNs to the System Operator. This will mean Suppliers will have to perform real balancing.
- Suppliers would have less of an influence in price setting than was envisaged because only a small proportion of demand is movable.
- Suppliers will be very exposed to the balancing market and imbalance price because they have very scant knowledge of their customer base and domestic demand is very variable.
- Suppliers will have a great incentive to work with customers, such as on demand reduction schemes. This will be useful for internal balancing, or to take advantage of balancing market decrements. Suppliers could offer customers favourable terms and conditions as a premium for demand management 'options' or as incentives for tailored profiles.
- Suppliers will have minimum bargaining leverage in long-term contract negotiations.

3.4.3 NETA's Losers

At present the primary function of Suppliers is to carry out the billing functions of their customers. Supply has a high proportion of fixed costs and has a very thin margin for error or mismanagement²⁹. Under NETA, Suppliers will have the responsibility for predicting their half-hourly demand, a new responsibility to this segment of the industry. They would be made to pay the penalty for getting their demand predictions wrong, as a result they would be faced with a much higher level of financial uncertainty for demand variability also known as volumetric risk.

As discussed in previous sections, in the desire to reduce their exposure to the shorter-term markets, Suppliers will face higher long-term contract prices than at present. Due to the nature of electricity consumption, knowledge of demand levels only becomes clearer closer to operating time. This would mean Suppliers would not be able to predict accurately and to cover their demand in the long-term. They will have to 'fine-tune' their demand in the short-term markets, exposing themselves to some degree of short-term price uncertainty.

Suppliers will have the steepest learning curve under NETA. They are expected to evolve from an industry with very limited knowledge on the behaviour of its customer base to a fully dynamic, and efficient trading machine. This makes the future for Suppliers very uncertain. The winners will be the ones capable of making the most effective transition in the shortest time. In conclusion, managing their exposure to the risks presented by NETA will be paramount to Suppliers survival.

3.5 Summary

NETA has resulted in a major overhaul of the ESI. The marginal pricing regime has been replaced by a series of bilateral markets. These bilateral markets operate on sequential bases starting with.

- Long-term bilateral contracts and exchange traded tools
- A short-term 'pay-as-bid' screen-based auction market operating from 24 hours until 4 hours before the scheduled half-hour;
- A balancing mechanism from 4hours until operating-time to allow the SO to balance the system;
- A settlement process for settling imbalances.

NETA's bilateral markets are expected to remove the market inefficiencies of the present system. However, the new markets could simply give Generators more leverage, without tackling the apparent cause of the failure of the marginal pricing system, the abuse of market power by large

portfolio Generators. This could cause a 'NETA effect', which could see increased prices, a rise in price volatility, a reduction in new entry, and an increase in long-term contract prices. NETA's most disenfranchised group will be Suppliers. Under the present system, they are the group with the tightest margins. The NETA effect could result in a further erosion of these margins through tighter imbalance penalties that will accrue from the scant knowledge of their customer demand habits and higher long-term contract costs.

4. Risk Management within the ESI

4.1 Overview

Suppliers are entering an increasingly uncertain era, with the introduction of NETA. They will need to evolve more effective risk management strategies to be able to stay competitive and remain in business in this new marketplace. This chapter examines the subject of risk management within the ESI, with a view to evolving a robust strategy for Suppliers to shield themselves from NETA's uncertainties. The definition of 'risk' is presented in the first section together with a general description of financial risk management.

Supplier risk management in the ESI is discussed in some detail, from the most commonly used techniques to the latest innovations. The inadequacies of these methods as they may fail to minimise risk under NETA are then analysed. This is followed by a proposed risk management strategy to eliminate Supplier short-term demand uncertainties under NETA

4.2 Definition of 'Risk'

Everything changes: change can be good or bad for those it affects. Change leads to risk or uncertainty, the prospect of gain or loss. The subject of risk is broad ranging and affects all aspect of everyone's daily lives, from crossing the road to eating food. For the purposes of this discussion financial risk as related to business operation shall only be considered. A suitable definition of financial risk is given as:

'the uncertainty in the outcome of the Profit and Loss statement'

Risk can thus be considered as the likely deviation from the expected outcome. Few business ventures or activities are entirely risk free. The three possible outcomes from a business project are:

- It makes as much money as was expected.
- It makes significantly less money than expected
- It makes significantly more money than expected.

The first outcome is the most desirable, because companies base their long-term investment strategies and dividend payout projections on expected profits or incomes. Realising these targets is vital to the long-term competitiveness and investor confidence in any business. The second outcome is undesirable and may lead to a loss if the revenue gained from the project is less than

the cost. In the worst case this may result in business failure. The final scenario is also undesirable, though not as disastrous. Unexpected revenue is often much less efficiently used because it is unaccounted for within the long-term business strategy. This could result in missed opportunity and may unduly raise investor expectation, which will have dire long-term consequences if the extra revenue, translated to higher dividend payout is not maintained.

4.2.1 Risk Management

Risk is inherent in all business activities and is something that successful competitive businesses must come to terms with. Coming to terms with the risk does not mean its elimination, which is clearly impossible; neither does it mean the fatalistic acceptance of its consequences. It means managing the risk; that is deciding which risks to avoid, and how to avoid them, which to take on and so on. The ‘science’ of managing risk is known as ‘risk management’.

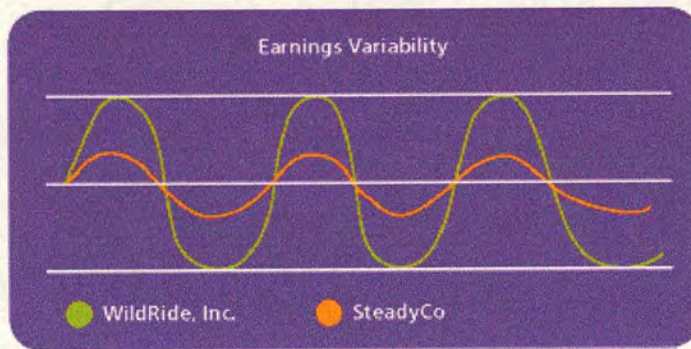


Figure 4.1. Example of earnings variability³⁰

An effective risk management strategy should ensure that companies have the cash for investment when needed. The strategy should not completely insulate the company from risk of all kinds, so as to be able to take advantage of favourable movements in fortune. Figure 4.1 shows the earnings variability of two companies, Wildride Inc., and SteadyCo. The income flows within SteadyCo are much more desirable because they are less variable and hence more predictable enabling its directors to plan more effectively than those in Wildride Inc.

4.3 General guidelines for the implementation of a risk management strategy

The first step to the formulation of an effective risk management philosophy is to impose guidelines on the risk management decision-making. This stipulates the kinds of risks to avoid and the ones to take on. The philosophy should give an indication of the attitude to take towards the various types of risks that might be faced.

Once a company decides to implement a risk management strategy, the following steps outlined below must be employed³¹.

- Identify potential risks: There are many sources of risk in a firm's operations and these must be identified
- Prioritisation: The potential risks must be ranked in order of their importance
- Risk analysis: Determine the nature of each risk, its effect on the company and the probability of the event occurring
- Decision: A decision is made as to whether the risks are trivial, tolerable or significant. This is used to rank the risks in order of urgency.
- Regulation: deciding on a course of action to be taken to eliminate or minimise the risk and analyse the effect of this action on the risk. As a further step, the effect of alternative risk management actions could be analysed and compared.
- Implementation: The enforcement and monitoring of the chosen course of action. It is possible that this could throw up other new potential hazards.

4.3.1 The risk of risk management

The benefits of applying risk analysis and management are great, but risk analysis and management are not without their own set of risks. The primary risk in doing any risk analysis is that the recommendations for the management of the risks are inaccurate. This is referred to as the risk of risk analysis. Understanding this risk is very important to ultimately increasing the value of the risk management strategy. The risks from risk analysis are outlined below³²:

- Inaccurate conclusion from risk analysis could either lead to overestimation or underestimation of the risks. Overestimation could have substantial effects like holding back assets. This could deprive key areas from the resource that could have made the business activities a greater success. Underestimation more often than not results in a secondary consequence, those of management panic. Instead of being proactive, management finds itself reacting to events.
- Perversely, accuracy can also be a cause of risk. The true value of the analysis cannot be proven, since by recommendation of the analysis, the risk was avoided. As nothing went wrong, the analysis can be seen as an unjustifiable cost. Thus, the next time the analysis is required it is not performed.
- Another risk with risk analysis is that it can be used to keep the status-quo. This can also take the form of organisational blame, causing intimidation instead of improving the situation.

- Risk analysis often over-relies on producing numbers, and does not rely enough upon the human expert knowledge and common sense required to interpret results. Different risk analysis techniques, produce varying results. This could result in arguments over the techniques rather than over what the results might be indicating.
- The cost of performing risk-analysis may impede project success. Firstly, risk analysis will take up resources that might be better applied to the actual project at hand. Secondly, risk is perceived as bad and the very fact of conducting an analysis could stigmatise the endeavour as one to be avoided.
- There are no quality controls on risk analysis themselves. Risks are only potential events with no certainties. Only after the event can the value of a risk management strategy be ascertained

An effective risk management strategy does not involve the identification of risks and the implementation of a course of action to manage these risks only, but must also feature an effective interpretation of the results and an understanding of the wider implications of the risk management strategy on the firm. Understanding the risks of risk management will ultimately improve the value of the strategy.

4.4 Quantifying Risk exposure

Before formulating a risk management strategy, the risk has first got to be quantified or measured. Risk measurement techniques are used for this purpose. Their aim is to provide an indication of the probabilities of different levels of deviation from the expected outcome. Financial risk measurement methods evolved initially in the financial markets. The earliest of these were Gap Analysis and Duration Analysis. Both were specifically tailored to measure interest rate risk. For all other financial risks, traditional measurement comes under two broad headings, statistical and scenario analysis. This section will briefly discuss both methods.

4.4.1 Statistical

Statistical analysis tries to postulate a measurable relationship between the exposure variable and the factors that might influence the loss or gain. These techniques are used to make numerical estimates of the uncertainty. The most commonly used technique in the financial markets is the variance since this gives a measure of the amount of spread of data about its mean. For example, returns on Wildride inc. have a higher variance than SteadyCo, therefore they would be considered more risky. A risk adverse investor will invest in SteadyCo to avoid the uncertainty. Variance has limited use as a measure of risk since it does not take into account expected

outcome. There exist more sophisticated techniques that attempt to provide an improved measure of risk, for example Pollatsek and Tversky³³. Their measure incorporates both variance and the value of the expected outcome, so that it can be more readily used to compare directly and to rank the desirability. Other knowledge about the company can also be gained from the variance calculation.

The disadvantage of statistical measures is that they are based on estimated parameters. These can be very sensitive to the choice of proxies and the inclusion or exclusion of other variables. Relationships can change over time, often quite dramatically, so there is always the added risk of under or overestimation because the relevant risk factor has altered.

4.4.2 Scenario Analysis

Scenario analysis gives a different type of measurement to anything obtained by statistical analysis. It estimates the effect of 'what if' scenarios on a company's activities. To perform this type of analysis, the spectrum of scenarios a business is likely to face are identified and stimulated to determine the impact on the business³⁴. There are two types of scenario analysis, static and dynamic. Scenario analysis can be applied to most kinds of risk, and is less limited by data availability and historical trends than statistical approaches.

There are four main steps in performing scenario analysis:

- Select a scenario, that is a path describing how relevant variables might evolve over a time horizon period;
- Postulate the cash flows and/or accounting values of assets and liabilities as they would develop under the scenario;
- Repeat the first two stages for any other scenarios that might be encountered;
- Use the results of the scenarios to come to some view about the exposure.

Scenario analysis is very effective because it alerts the company to all possible outcomes from a business endeavour. When using this method, care should be taken to explore all possible scenarios, as the omission of any potential scenarios will mean risks could go undetected.

4.5 Risk management methods – hedging

Once the exposure has been determined, a decision must be made based on the corporate risk philosophy, on how to manage this exposure. If a company decides to take a proactive approach towards managing the risk, then the concept of hedging will be utilised. Hedging is the most commonly used risk management technique. A hedge is a contract entered into or an asset held as a protection against possible financial loss and will function in either one of two ways.

- If the exposed asset (or liability) losses value, the hedge will compensate by increasing in value.
- If the exposed asset (or liability) gains value, the hedge balances by decreasing in value.

Hedging reduces the volatility of a company's cash flows, giving greater certainty to future cash flows and consequently enhanced budgetary decisions. There are two types of hedges, the internal and external hedge³⁵.

- An internal hedge is constructed by matching the risks with a high negative correlation within a company's portfolio. For example, portfolio generators own a diversity of plants with different fuel and generating characteristics. This diversity reduces exposure to plant specific risk.
- The external hedge involves offsetting the financial risk with an external party willing to take on the risk either for a fee or to balance an equal but opposite risk within their portfolio. This external agent is known as a counter party. The external hedge is done through either purchasing exchange traded or Over-The-Counter (OTC) instruments. OTC instruments are tailor-made to fit exactly the hedging requirements but they are rather expensive and involve an element of counterparty default risk.

4.5.1 Hedging strategies

There are two main types of hedging strategies, static and dynamic. Static hedging involves taking a fixed position for the duration of the hedge and dynamic hedging involves actively changing the hedging position to take advantage of more favourable opportunities during the lifetime of the hedge.

Once the decision has been made to hedge, the basic choice is between standardised exchange-traded instruments and tailor made OTC instruments. The trade-off is between exchange traded instruments which are more liquid, and OTC which fit hedging needs more closely but are more expensive and less liquid.

To illustrate the concept of hedging, an example of a simple static hedge using exchange-traded options is presented. PipesCo is a large copper pipes manufacturer that will start purchasing a significant quantity of copper for a large order in 6 months time. Assuming that there is no exchange rate risk, then PipesCo will be exposed to the spot price of copper on the commodities markets in six months time. The company will favour lower than anticipated prices in the future because higher prices will increase production costs. PipesCo will want to insulate themselves

from higher prices while having the option to take advantage of lower prices. To achieve this, PipeCo will buy 'call' options on the future price of copper. A call option gives the holder the right to buy the asset at an agreed price on a specified date. A profit diagram for this example is illustrated in figure 4.2.

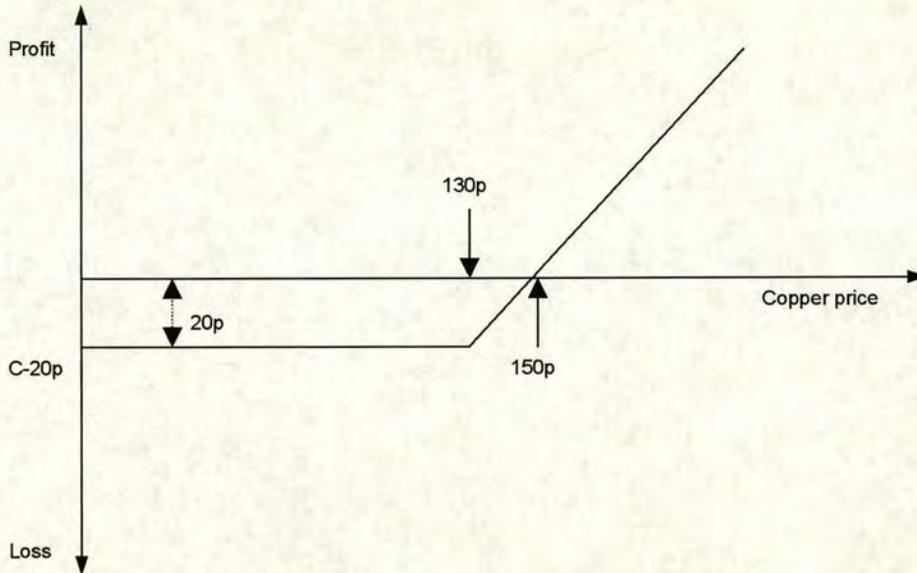


Figure 4.2. Profit diagram for 'call' option of copper prices

The agreed future price on the commodity market is 130p per unit of copper (known as the exercise price) and the specified date for delivery is in six months time (the expiration date). If the price of copper is less than the exercise price on the expiry date, the option will be allowed to expire, costing PipesCo 20p, the price of the option. If the copper price is above the exercise price, then the option is exercised and PipesCo will pay 130p for each unit of copper plus 20p, the price of each option. The option could be viewed as an insurance policy since the fee paid is the premium to guarantee a cap on the maximum price for a unit of copper. By pursuing this strategy PipesCo has ensured predictable future cash flows.

4.6 Overview of risks in the Electricity Supply Industry

Up to this point, the discussion has focused in general terms to introduce the concepts of risk and its management as can be applied to any industry.

The subject of risk management has been relevant in the ESI since its inception mainly because of the dynamic nature of electricity supply. The risk management in the pre-privatisation era was aimed primarily at optimising generation, transmission and supply. Post-privatisation risk management has been very different, and is geared towards increasing shareholder value. This

has led to the evolution of a new generation of methods and tools, the majority of which were borrowed and adopted from the financial markets.

This section focuses specifically on risk and its management by Suppliers. These risks will be the same regardless of the market structure but their manifestations might vary.

Supplier's financial risk can be grouped under the following headings; credit, volume, price, basis and operational risks. Particular emphasis is placed on volume, price and basis risks because they represent the uncertainty from Suppliers core business activity, supplying electricity.

4.6.1 Credit risk

Credit risk is the risk that a loss will be incurred if a counterparty defaults or fails to meet the terms on any contract. There are two types of credit risk, settlement and performance risk. Supplier settlement risk is associated with the anticipated payment for power delivered. It occurs when either a Generator fails to honour the output commitments stipulated within the terms of a long-term contract or a customer defaults on their bill repayments.

Performance risk is the risk resulting from the difference between the market price and the contract price of power delivered at an agreed price. In the wholesale markets, Suppliers sign fixed priced contracts with Generators. If the long-term average price of electricity in the spot market is lower than the contract reference price, then the extra revenue could be viewed as a missed opportunity.

So far, settlement credit risk has not been a problem in the UK electricity industry. However, the gas market has seen some bankruptcies from settlement risk and also, the US power market suffered some major defaults in June 1998³⁶. The ESI's immunity from settlement credit risk has been because most of the market participants are fairly large and are cash rich. Additionally, the Pool acted as counterparty to all trades. In the unlikely case of a default, the Pool makes up the deficit, shielding Suppliers from the brunt of this risk.

Under NETA, companies will trade bilaterally and there will be no pool to act as counterparty to all trades. Participants will thus be exposed to higher levels of settlement risk. Also, greater market uncertainties will translate to income uncertainty. Thus increasing the default risk of more market participants.

4.6.2 Operational risk

Operational risk exists within poorly managed or structured organisations. This is the risk that the company would run into trouble through badly managed trading activities. It is a risk that affects all companies.

4.6.3 Basis risk

Basis risk occurs where the cashflows from the financial tool used to hedge do not match those of the risk they are intended to hedge. An example is when a Supplier buys an Electricity Forward Agreement (EFA) to hedge against future Pool price. EFA's are swaps where the buyer and seller agree a forward Pool price for a fixed quantity of electricity. Closer to the future delivery date, the buyer's demand circumstances change negating the need for the electricity purchased through the EFA. In this case the seller will be left with unwanted volume. The risk of this type of cash flow mismatch is known as basis risk. In more liquid markets the mismatch is not a big problem because mismatched volume can be easily liquidated. This is not the case in the electricity markets, because the low volume of trade in the hedging products, EFA's and CFD's, would make it difficult to unwind positions. Under these circumstances, such mismatches could have serious financial implications³⁷.

The principal types of mismatch are:

- **Product Basis:** This is the most important basis risk in the energy markets. It arises when there is a mismatch in quality, consistency, weight, or other specification of the hedge and the underlying product. Product basis is a big problem because positions as in the example above are not easily liquidated.
- **Time Basis:** This risk occurs when there is a sudden shift in demand or transportation bottlenecks occur. An example of a Supplier hedging its winter position by purchasing January swing options on contracts is used to illustrate time basis. For a fee, swing options give the Supplier the option to vary the amount of electricity purchased from the stipulated contract quantity. If the severe cold spell were to arrive early in the winter, say in late November, then the spot market prices in December may surge much more than the January prices. In this circumstance, the January swing option will not provide an efficient hedge against December's spot price and physical requirement
- **Locational Basis :** This risk was not a major problem because the Pool provided a uniform price to Suppliers, regardless of location. The story is entirely different under NETA's balancing market where the system will be balanced at the various Bulk Supply Points (BSP) around the country, and balance prices will be determined according to the

constraints at each point. This will result in differential locational Balancing Market prices. Suppliers will have to absorb these price differences, because they will be obliged to charge uniformly across the country.

- **Mixed Basis :** Another type of basis risk in the energy markets is mixed basis. It occurs when an underlying position is hedged with more than one type of mismatch. A Supplier might hedge a January demand exposure with a February gas future, leaving both time and product basis exposures. Gas futures are sometimes used to hedge electricity exposures because of the intrinsic link between both markets.

4.6.4 Price risk

Pool price risk occurs whenever the pool price varies from its expected value. The pool price varies considerably on a day-to-day basis. This is illustrated in Figure 4.3 which shows the half-hourly daily pool prices from three days in the same week; Sunday 8th October 2000; Wednesday 11th October 2000; and Thursday 12th October 2000. There are variations in prices within the day as well as between days. The high level of variation makes the Pool price very hard to predict, exposing participants to income uncertainties.

Pool price is very volatile hence difficult to predict because it is determined by a complex array of interdependencies. Electricity is difficult to store, as such supply and demand have to be constantly matched in real-time. The external influences of gas market arbitrage opportunities and market power within the price determination process increase its complexity.

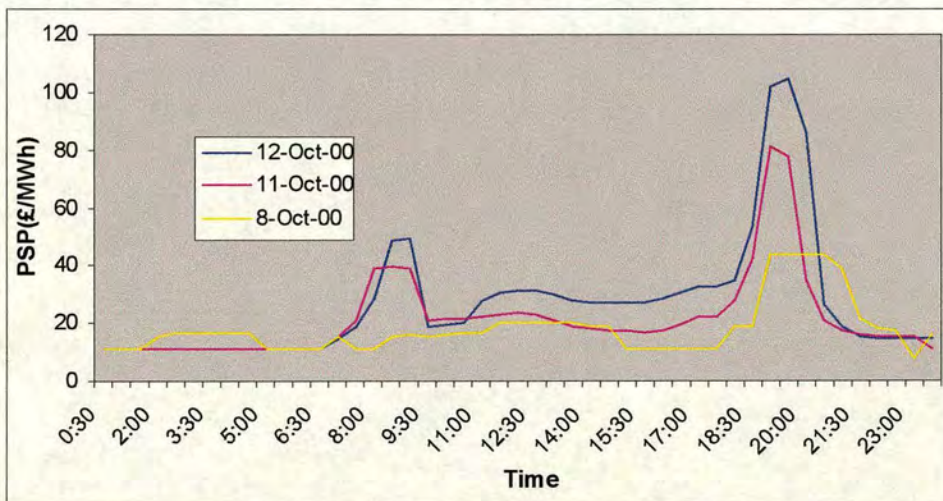


Figure 4.3. Half-hourly pool price from 3 days in the second week of October 2000

Volatility is the degree of randomness within the pool price. Figure 4.4 shows the historical half-hourly pool price volatility computed over four years from 1996 to 2000. The volatility peak occurs between 4:30 and 8pm, the evening peak, this coincides with the daily price peak. The combination of the highest prices and higher volatility causes high price exposures over this period. Volatility is particularly high during winter, when ratios of peak prices to prices during the rest of the day of 10:1 are not uncommon.

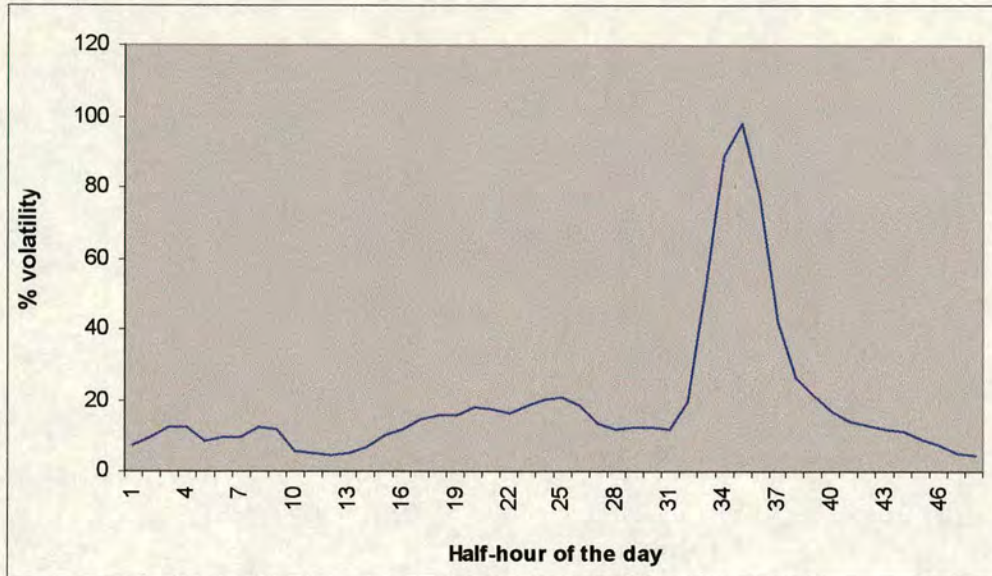


Figure 4.4. Historical volatility of Pool Selling Price for period 15/11/95 – 8/8/99

The capacity mechanism within the Pool Selling Price has served to dampen the volatility at the supply and demand margins. There is no such mechanism under NETA. In bilateral markets supply and demand are matched closer to operating time as such volatility will be much higher. NETA's volatility might be similar to other bilateral electricity markets without capacity incentives. For example in the Victoria Pool in Australia, prices for the same half hour of the day ranged from A\$0 to A\$5000 (US\$2899) during 1998³⁷. This anticipated price volatility increase will expose Suppliers to greater electricity price risk.

4.6.5 Volumetric Risk

Volumetric risk is when the actual demand varies from expected demand. It is caused by customer deviations from their expected consumption pattern. Electricity end-users are classified into three generic groups, industrial, commercial and domestic, based on their level of consumption. The industrial load segment tends to be fairly predictable because the majority of their consumption is used on industrial processes. Commercial and Domestic loads tend to be more variable because they are heavily dependent on the weather and random influences like the World Cup or Princess Diana's funeral.

Suppliers buy fixed volume contracts for greater than 80% of their expected demand in advance. Volumetric risk is less of a problem if the half-hourly demand is less than the contracted volume. However, if the demand exceeds the Suppliers contracted volume, the Supplier will have to make up the excess demand on the spot market. This will expose the Supplier to NETA's imbalance prices.

There are two types of volumetric risk: linear or parallel shift, and non-linear or slope and curvature shifts in the load profile:

- Parallel shift: Each daily half hourly load increases by the same amount, this causes the whole daily load shape to move up and down
- Non-linear shift: The half hourly load during a day changes in different proportion, this changes the shape of the daily load curve. For example, if a customer takes more power from 1pm to 6pm in a day and keeps the same consumption level of power for remainder of the day, then both slope and curvature in the load shape has changed.

Non-linear shift is very hard to hedge against because the load shape changes: parallel shift is less problematic because the hedge can be tailored to fit the load shape. This can be lower or raised to fit the actual demand shape.

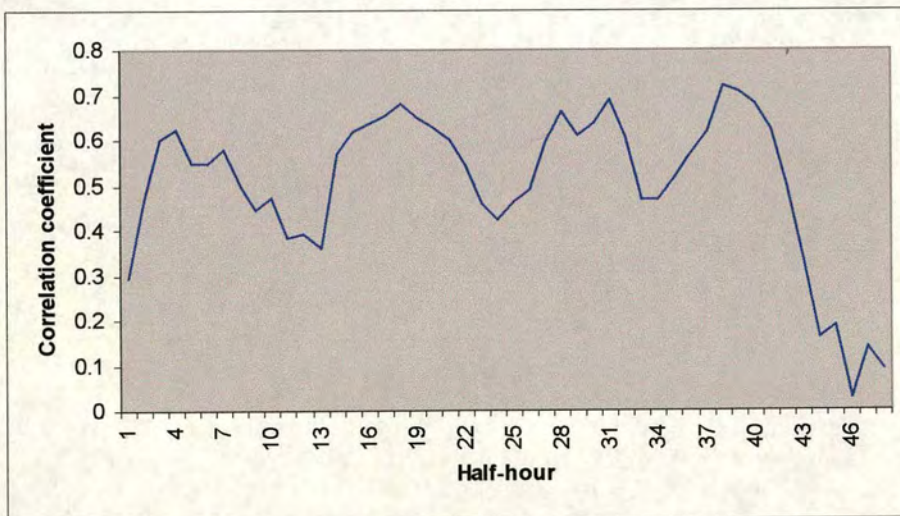


Figure 4.5 Correlation of pool price versus time of day at a local 33kV BSP for demand in 1997

The financial implication of volumetric risk becomes greater the higher the positive correlation between spot price and demand. Conversely, negative demand and price correlation implies

Suppliers could buy the extra volume at lower prices and could possibly benefit from prices lower than the contract price. An example of price and demand correlation is shown in figure 4.5. It illustrates the correlation between the price and a predominantly domestic load, metered at a 33kV Bulk Supply Point (BSP) in Scotland for 1997. There is a varying positive correlation between the price and demand throughout the day. There are four demand peaks at this location; the first peak between 2am and 6am, is due to the overnight heating load; the second peak is due to the morning demand; the third peak is the afternoon heating load; and the fourth and highest peak between 5pm and 8pm represents the evening demand pick-up.

In the UK, with the exception of larger half-hourly metered customers, all end-users are charged a fixed rate based on fitting their aggregated metered demand to a set of pre-defined standard daily profiles. As mentioned earlier, the smaller customers tend to present the greatest demand variation. This demand segment does not pay for their electricity according to the specific time of day it was consumed. The use of standard daily profiles passes on the cost of volumetric risk directly to the Supplier. Standard daily profiles give very little incentive to customers to reduce their time of day demand variability.

Though volumetric risk will stay the same, its financial implications will increase with the introduction of NETA. This is firstly because the bilateral market prices will be more reflective of supply and demand conditions, resulting in an even higher correlation between electricity price and volume. Secondly, the price volatility will increase resulting in higher price uncertainty. Thirdly, there will be stiff penalties imposed on out-of-balance participants.

4.6.6 Summary of possible evolution of risks under NETA

As at the time of writing the risks under NETA are not fully known, they are expected to become apparent sometime after its inception. This section presents a summary of the possible evolution of the risks outlined in the previous section, based on the point of view held both by the author and many observers³⁸.

- Credit risk: This is expected to increase because the risk of default will be greater as participants are expected to experience more volatile income flows.
- Operational risk: This is expected to increase in the early days of NETA as companies adjust their systems to the new trading arrangements.
- Basis risk: There will be a wider array of instruments to trade electricity. This will increase the scope for mismatches, potentially increasing the Product basis risk. Time basis should stay the same as it is at present. Locational basis could become a big problem if balancing mechanism prices reflect local constraints.

- Price risk: It is expected that prices in the various markets will be more volatile than under the Pool, resulting in greater price risk.
- Volumetric risk: Though the physical level of volumetric risk should stay the same, the financial implications will be increased.

4.7 Risk measurement and management in the ESI

The risks in the ESI were outlined in the previous section, and the expected evolution of these risks under NETA was summarised. The first step towards effective risk management is risk quantification. This section will examine the risk measurement techniques used within the Pool and discuss their inadequacies and possible modifications to suit NETAs needs. The methods used were based on techniques used by Scottish Hydro-Electric(SHE): quantification of price and volumetric risks are discussed. Credit risk is not discussed in this section because it was not a problem under the previous system so risk measurement and management was not necessary.

4.7.1 Determining the level of exposure to Pool price risk

The level of exposure to Pool price risk was directly proportional to the percentage of the Supplier's contracted volume. Lower levels of contracting meant the Supplier must make up the excess volume on the spot market, the Pool. Contracts were hedges against Pool price risk and as such, the higher the percentage of contracting, the lower the Pool price exposure. The method used by SHE to determine Pool price risk was fairly standard, and is presented below as a series of steps to show a typical calculation.

- Exposure was calculated as the percentage of uncontracted demand on a 3-month rolling average
- Forecasts of the pool price were obtained from the National Grid Company (NGC) these forecasts were made from historical prices, and knowledge of future supply and demand
- The average error of previous forecasts for the month in question was calculated by subtracting the actual price from the historic price of the particular month for the previous three years to give the Mean Absolute Percentage Error (MAPE) of the forecasts
- The Pool price exposure was calculated as the product of MAPE, net contracted volume and forecast price

A risk management team was responsible for keeping the exposure within the limits set by the Board of Directors in their monthly budgets. The team might choose to increase or reduce the level of participation on the spot market based on the exposure calculation. During the winter months when the demand was generally higher and prices more volatile, the risk manager opted for a greater level of cover than in the summer months. From previous experience, the level of exposure to the spot market was generally less than 20% of the demand⁴⁰.

The perception of future spot price levels and volatility determined the level of contracting required by Suppliers. This gave Generators the negotiating power to charge more for contract cover. Suppliers effectively paid a risk premium for higher contracting.

4.7.1.1 Inadequacy under NETA

The nature of long-term price trends was not known at the time of writing. Additionally, the single reference price, Pool price, was to be replaced by a basket of prices. The method detailed above would require modification before it can be applied to the new markets. It is very important that participants do not rush into quick fix solutions to manage price risk because the true nature of prices and volatility will only be understood after a long period of observation.

4.7.2 Determining the level of volumetric risk

To determine volumetric risk, customers were grouped into two classes. The first group were the larger wholesale industrial and commercial customers who signed tailor-made contracts with Suppliers for their electricity. The second group were the smaller commercial and domestic customers, who were charged based on standard tariffs.

4.7.2.1 Larger customers

Fixed price contracts were usually signed with the larger customer group for at least six months to a year in advance. Such contracts might stipulate a single tariff for a unit of electricity, e.g. £15 per MWh, or Standard Time of Day (SToD) tariffs to reflect the higher cost of electricity during peak periods and the lower costs at times of low demand. Tariffs were normally calculated from historical records to recoup the expected cost of electricity the customer might consume. The most commonly used approach for determining customers volumetric risk, was the load factor. It is a dimensionless coefficient used to determine the level of demand variability.

The load factor was calculated as:

$$\text{Load factor} = \text{Average Demand} / \text{Maximum Demand}$$

This measure indicated roughly how 'peaky' a customer's expected demand was. Lower Load factors tended to indicate high variability during the day and/or highly seasonal load. Peak customers were riskier to Suppliers because their demand was more difficult to hedge. As a result Suppliers were exposed to price risk, because the excess demand was purchased from the spot market. High Load Factors implied flatter demand, which implied less variability, which was easier and less expensive to hedge.

Generally, customers with higher load factors were more desirable because they presented less risk to the supplier. Although Load factor was not ideal, it is the most commonly used risk measure within the industry because it is widely understood⁴¹.

A typical load profile for a large customer is shown in Figure 4.6. As can be seen, this customer has a flat demand for electricity with two peaks after midnight. This customer used electricity overnight and had negotiated cheaper tariffs with his Supplier. This customer was very desirable because his demand was predictable and was flat during peak periods: the customer's load factor is 82.18%.

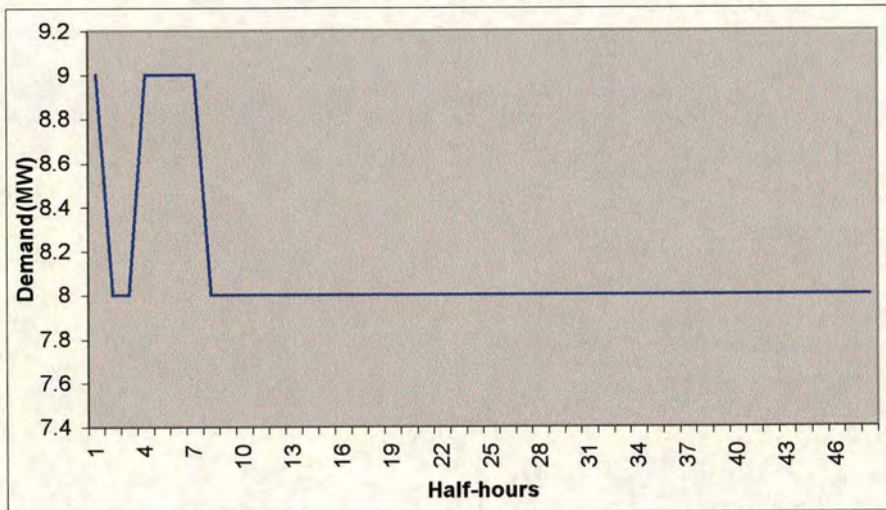


Figure 4.6. Industrial customer demand on 5th January 1998

In practice, the use of more creative tariffing reflective of time-of-use, was rather limited because larger customers preferred simpler, more easily understood tariffs. For this customer segment, there was stiff competition between Suppliers and electricity expenditure only made up a small percentage of the customer's budgeted expenditure. Larger customer's demand preferences tended to dictate the style of contracts Suppliers issued. As a result, ambitious tariffing schemes tended to be abandoned in favour of simpler ones.

Empirical studies show that customer responses to STOD are often very limited. Spector, Tishler and Yingu⁴² found that customers failed to change their pattern of output significantly regardless of the potential savings. The reason is that customers' perceived adjusted costs, for example scheduling work shifts to avoid peak price, outweighed the perceived benefits of a shift in demand.

4.7.2.2 Smaller customers

Smaller customers were charged according to their aggregated demand over a time period, superimposed onto a predefined set of load profiles. The load profiles were designed by the Electricity Association and are based on statistical data collected over time from monitoring a wide sample of users⁴³. There are eight generic classes for the under 100kW customer group, two domestic and six non-domestic.

This customer segment causes the greatest variability to the Supplier's demand. Though profiling is more reflective of time-of-use pricing than single tariff pricing, it does not reflect individual customer variability because it assumes uniformity across a generic group. There is no method used at present by Suppliers to determine the level of risk from this customer segment. For the recovery of individual customer costs, more reflective of time-of-use, chunking and algorithmic profiling were introduced⁴⁴

Algorithmic profiling stretches or shrinks the consumption duration on a standard load profile whilst maintaining the overall daily consumption. This is to accommodate customers on demand-managed heating loads. Chunking is when the daily load profile is either aggregated upwards or downwards based on the total quarterly metered demand. Algorithmic profiling and chunking assume parallel shift risk. Though they are better alternatives to standard profiling neither method accounts for non-linear shift risk, as the modified profiles are not reflective of individual demand variations.

Load factor calculations could also be used to determine volumetric risk in this segment. There is no half-hourly metering in this demand segment, so volumetric risk was determined from measures at Bulk supply Points (BSPs). If there were large customers within the BSP group, their demand was subtracted from the aggregate demand. The load factor calculation is illustrated with an example from a predominantly domestic BSP in northwest Scotland between 1997 and 1998. The load factors in both years for the month of January were calculated as 64.53% and 64.85%. This showed very little change across the years indicating that the demand profile in this section was fairly fixed. In this case, the customer-specific random influences seem to be cancelling out each other. The load factors for both years were 55.63% and 57.46%. This was lower than for the month of January, indicating that the yearly profile was more variable than the within month profile. This was to be expected because of seasonality. Load factor is a very blunt tool and is not an exact measure of load variability. Figures 4.6a and 4.6b show the daily demands from the BSP for week 3 in January 1997 and 1998. From the graphs, the similarity of the daily profiles across week 3 in 1997 and 1998 vindicate the load factor calculation.

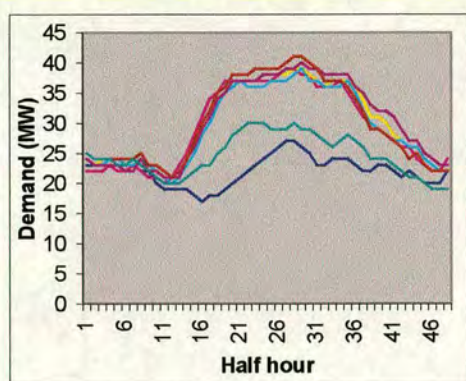


Figure 4.6a Week3 1997

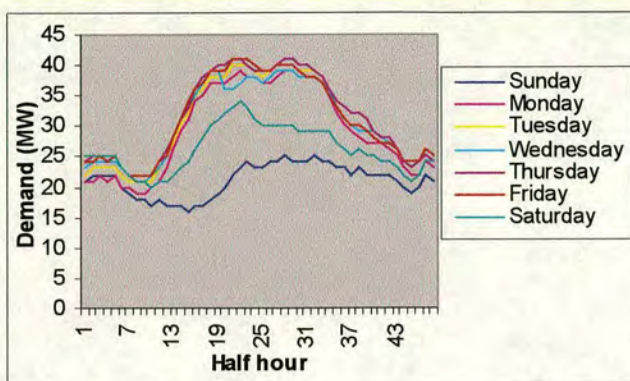


Figure 4.6b Week3 1998

4.7.2.3 Application of volumetric risk management tools under NETA

As previously mentioned the financial implication of volumetric risk is much greater under NETA. The risk measurement tools described above are applicable with the new arrangements, however these methods are fairly blunt and should be used in conjunction with more sophisticated methods such as forecasting to offer effective risk measurement.

4.8 Risk management tools

So far the risks within the ESI have been presented, followed by the methods used to quantify these risks. This section examines the tools used to manage these risks, once the exposure has been determined, or the decision has been taken to hedge.

4.8.1 The CFD Market

The price risk caused by Pool price volatility was hedged through the use of over-the-counter (OTC) derivative instruments, namely Contracts For Difference (CFD). These were born out of a necessity to hedge all generator output at privatisation. To guarantee the purchase of their output, bilateral contracts were signed with the RECs. In 1999, 80% of the total demand in the UK was traded through CFD's⁴⁵. At the time of writing, the industry was in the process of modifying these contracts to meet NETA's requirements.

CFDs typically had a one to two year term, but it was not uncommon for contracts to be signed for longer periods. Longer-term contracts were usually linked to long-term fuel purchase arrangements. For example, 15 year CFD's were used to link Independent Power Producers (IPP) sales revenue to fuel purchase and financing arrangements.

CFDs were essentially financial swaps between Generators and Suppliers. There were two generic classes of CFD's, stacking contracts and time dependent contracts. These are expected to undergo some evolution under NETA however their basic principles should stay the same.

4.8.1.1 Stacking contracts

For a fixed fee, the seller of the contract agreed to pay the buyer the difference between the Pool price and contract exercise price for each half-hour the Pool price exceeded the exercise price. The exercise price was the price above or below which the contract became active. For the case of two-way CFDs, the buyer also agreed to pay the seller the difference between the Pool price and the exercise price whenever the Pool price was below the exercise price. The half-hours covered by the contract were stacked in descending order of price. The two-way CFD is illustrated in figure 4.7. Under NETA the pool price might be replaced by an index made up of a basket of the prices in the various markets.

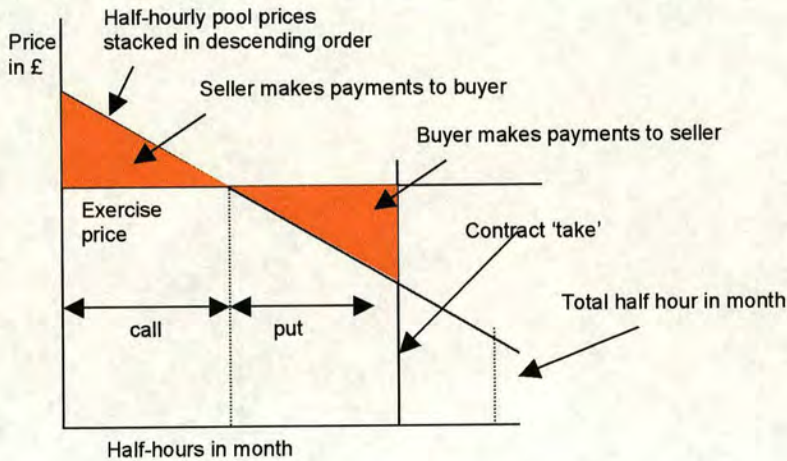


Figure 4.7. Two-way stacking CFD contract

4.8.1.2 Time-dependent contracts

These were more popular because most Generators and Suppliers found it easier to predict their output and demand with reference to time rather than to the level of Pool price. Time related contracts were the most commonly used CFD type. It is expected that these types of contracts would be in higher demand under NETA because prices are expected to be more volatile and thus harder to predict than Pool price.

The contract was usually structured as a two-way option, with difference payments being made either way depending on the Pool and exercise price. In this fashion the contract effectively swapped the Pool price for the contract price. Contract quantities varied according to the season and the time of the day. The periods covered in the contract were based on any standard time interval, for example time-of-day, day of week or season, or might have been sculpted to reflect changes in the level of cover required. An example of a time related contract is given below in figure 4.8. The contract is simply a function the time of day.

CFD's effectively swapped Pool price for the contract price. Contracts generally used PPP, which was the sum of the energy (SMP) and capacity components of Pool price, as the reference because it was the common risk shared by Generators and Suppliers. CFDs on PSP were sometimes also designed. The relationship between the contract price and exercise price was arbitrary; it reflected the risk premium that was judged fair by both parties.

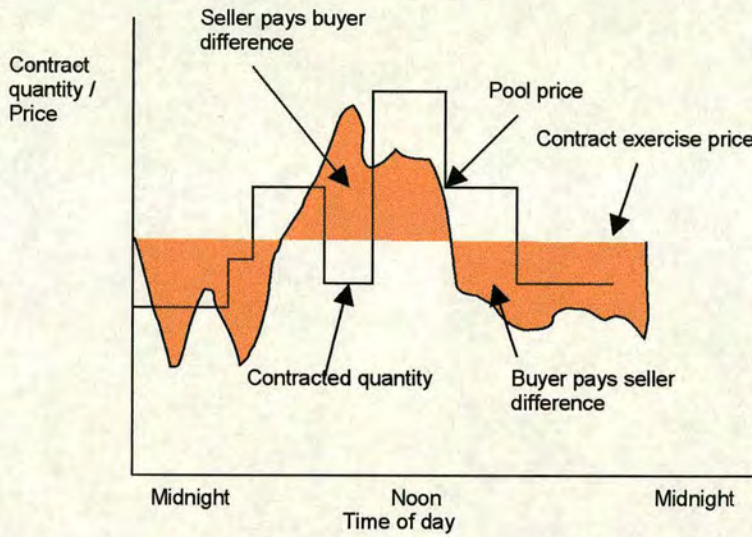


Figure 4.8. Time-dependent CFD contract

As with Stacking contracts, it is expected that a basket index of NETA prices would replace the Pool price as the reference for these contracts.

4.8.2 The Electricity Forward Agreement (EFA)

Shortly after privatisation, the industry realised that CFDs were too large and inflexible to meet all the hedging needs of participants. CFDs were purely bilateral agreements and, as a result the larger Generators used their market power to influence contract prices, making the market very opaque. The CFDs market was also biased against the smaller participants and deterred outside participation, hampering liquidity. EFA's were introduced in 1991 as standardised instruments that were supposed to give more hedging flexibility and encourage outside participation, thus enhancing liquidity. Although EFAs were initially traded through a third party, in time most agreements were made bilaterally.

EFAs were basically swaps. They were 2-way forward contracts that do not involve payment of an up-front fee. The buyer and seller agreed a forward price for the Pool over a defined period in the future. If the price turned out to be greater than the agreed price, the seller compensated the buyer. If it turned out to be less, the buyer compensated the seller. The price reference used for

the majority of EFA’s contracts was PPP: the remainder of transactions used capacity or SMP as their reference.

In the EFA market, the year was divided into calendar weeks, then, for each day forming part of a week, into six four hour slots. The four-hour period was chosen to smooth out any short-term spikes in the Pool price, while maintaining hedging effectiveness, and condensing market activity into manageable series of trading blocks. Table 4.1 below shows the EFA time periods. Weekdays are treated differently to weekends because of their different demand patterns. The minimum cover period that could be purchased was a 4 hour period for each weekday (20 hours) or weekend (8 hours) in units of at least 1MW.

	Mon-Fri	Sat+Sun
23:00-03:00	WD1	WE1
03:00-07:00	WD2	WE2
07:00-11:00	WD3	WE3
11:00-15:00	WD4	WE4
15:00-19:00	WD5	WE5
19:00-23:00	WD6	WE6

Table 4. 1. Time periods in the EFA market

EFA periods were seldom traded in isolation and were viewed as a series of standardised ‘building blocks’ with which market participants might sculpt a contract based on their hedging needs. For example WD3, WD4, and WD5 could be combined to provide weekday daylight cover over a month, winter or a whole year. A few contract structures became more popular and were more frequently quoted. Evident among these was Loadshape 44, which was particularly popular with Suppliers. It was an annual swap, which provided a volume of 40MW on weekdays between 07:00 and 19:00 hours, with half that volume during the remainder of the week. Loadshape 44 for weekdays and weekends is illustrated in figures 4.9a and b. This shape was subdivided into seasons of six or three months and was particularly useful to Suppliers when managing their portfolios.

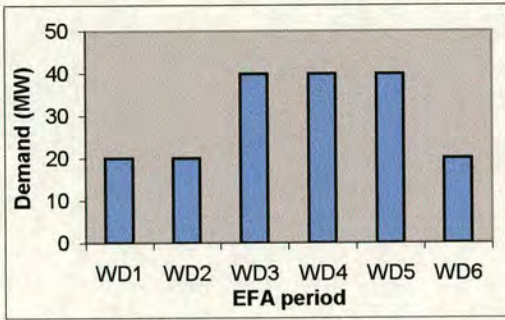


Figure 4.9a. Weekday loadshape 44

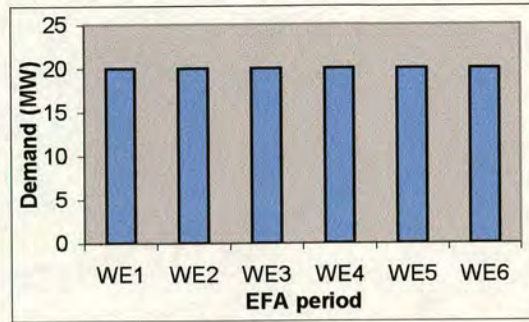


Figure 4.9b. Weekend loadshape44

EFAs were usually used in tandem with CFDs. CFDs were purchased for cover over the medium term (one year). During this time period the market might take temporary deviations from the forecast on which the CFD hedge was based. EFAs were purchased to provide temporary cover over such periods. In other cases the forecast might deem hedging cover either unnecessary, or temporary blips could arouse the need for short-term cover. EFAs were best suited for this purpose. They provided greater flexibility affording more exotic hedging possibilities and were cheaper than CFDs. It is envisaged that EFAs will also still have a major risk management role to play under NETA. Like CFDs, the contract price reference will change.

4.8.3 Swing options

Some CFD contracts, for a fee, had options allowing Suppliers to vary the quantities of electricity demanded from the stipulated contract amount. These were known as swing options. There were various types of swing options in the marketplace. However, the two main distinct swing option groups, based on the types of counterparties in question are: demand and price driven swing options⁴⁶. These are discussed in turn.

4.8.3.1 Price-driven swing options

Price-driven swing options allowed counterparties, the Generator and Supplier, to buy and sell quantities in the market place. When the electricity price was high the Supplier could nominate the contract upwards and sell the extra volume to the Pool. Conversely, when the Pool price was very low, the Supplier could nominate the contract downwards and purchase his demand commitments from the Pool. This type of contract allowed the Supplier to maximise the swing contract value through the purchase and sale of energy. This option took on various shapes and forms; the basic contract allowed the base load of energy delivered to swing a certain amount. Daily and monthly maximum and minimum quantities were defined at the start. Another variation limited the option holder to the amount of times the quantity was allowed to swing from the base load quantity. Multiple peaking options had no base load quantity and allowed the Supplier to purchase the same quantity of energy but for a fixed number of days over a defined

time period. For example, the option on contracts for next day delivery might stipulate that it could be exercised five times only during the summer.

4.8.3.2 Demand-driven swing options.

Demand swing options were found in contracts where Suppliers could nominate only up or down the quantity demanded and hence they could only accept and not sell the energy into the market. Suppliers nominated the swing contracts to meet increased or reduced demand conditions. For an up-front fixed fee, some demand-driven options were allowed to swing without bounds, that is, without limits to the quantities demanded. Others had the maximum and minimum quantities defined.

Swing options will have an important role under NETA. They can be used in conjunction with CFDs and EFAs to offer Suppliers the flexibility to alter their contracted demand positions closer to operating-time to avoid imbalance charges.

4.8.4 Weather derivatives

Since their introduction in 1997 by Natsource (a risk management company), latest estimates put the number of weather linked deals done in the US by 2000 at between 2500 and 3000, with a total value of around \$7billion. In Europe, around 100 deals with a total value of \$45 million were struck in 2000. The rate of growth is particularly notable in Europe where only a handful of transactions were conducted in 1999⁴⁷. Weather derivatives are OTC contracts that swap the risk of unfavourable weather on company returns, through a third party, with a company that has equal but opposite weather risks. For example, an ice-cream manufacturer could enter into a summer weather derivative with an electricity Supplier. The ice-cream manufacturer is likely to be exposed to low sales volume during a poor summer because demand for ice cream may decrease. This is the opposite for an electricity Supplier who benefits from increased consumption when the weather is bad. The ice-cream manufacturer would compensate the Supplier in the event of a warm summer and will receive compensation from the Supplier in the event of a poor summer.

The weather is the main source of customer demand variation and, the primary cause of volumetric risk. It causes domestic consumers to consume more by staying in and using higher levels of space and water heating. Weather-induced demand fluctuations can swing volumes by 30% from one year to the next. Instruments to manage this risk will be very important under NETA. At present there has been very limited use of the appropriate instrument among Suppliers primarily due to scepticism. The only publicised deal was between Scottish Hydro-Electric (SHE) and Enron, a major US energy dealer⁴⁸. SHE was compensated whenever the average temperature remained below a specified level during the winter 1998/99, through a Heating Degree-Day (HDD) swap, lasting from the beginning of October to the end of March.

The purchaser of a HDD swap receives a certain value for each unit of temperature below a set reference temperature on each day for the life of the swap. A degree-day is the number of units of temperature below the reference temperature on a given day. In the SHE/Enron deal, the location reference point was set somewhere in England. Due to the lack of data on this deal because of commercial sensitivity, a reference point in North Eastern Scotland would be used to illustrate a similar weather derivative contract. The reference price for the contract is 65°F : the same as was used in the SHE swap. The graph in figure 4.10 shows the average daily temperature and the reference temperature threshold line. The temperature data was purchased from the Meterological Office for the Dyce weather station near the Aberdeen airport. The sum of money agreed for each degree-day is £5, which is multiplied by the sum of the HDDs for the duration of the contract. This gives a total of 4350-degree days. For a winter swap between October 1998 and March 1999, this contract is due £21750.

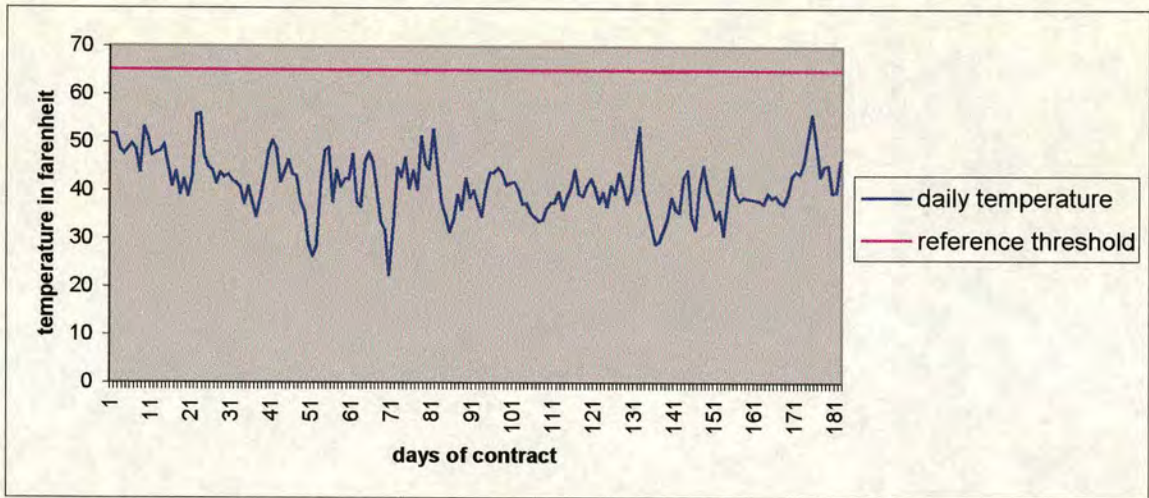


Figure 4.10. Weather derivative contract

The HDD calculation illustrated above is the simplest valuation method for weather derivatives and was the method used in the Scottish-Hydro swap.

An estimated 70% of all businesses face weather risk in some form and as such weather derivatives will be of use to most companies. The weather derivative market is still mostly an OTC market. In Europe, the high cost of acquiring weather data, in some cases adding 10% onto the cost of the contract, has also been a major factor slowing their uptake. From the evidence of growth over the past three years, it is expected that initial scepticism will thaw and weather derivatives will become even more popular over the next decade, leading to increased liquidity and more transparent and flexible products. They could to play an increasingly important part in Suppliers risk management activities under NETA.

4.8.5 Demand Side Management

Demand side management (DSM) was originally set up in the nationalised era to decrease end-user electricity demand in order to cut overall demand growth. This was done to reduce the need for further investment in generation and distribution plant. Although, in the search for profit, it is often in the interest of Suppliers to sell more electricity rather than less, sometimes it is advantageous for the utility to help reduce its end-user demand. For example, a Supplier would want to minimise his demand during periods when electricity costs to him are at peak values. On occasion, the avoided cost to meet the greater demand is less than the potential profit from greater sales. This is known as Least Cost Planning (LCP)⁴⁹. The post-privatised vertically industrial structure of separate generation, transmission, distribution and supply, make such optimisations difficult to achieve. In Scotland, where vertical integration has lasted for much longer, there is greater demand-side management.

Despite these limitations, DSM might be very useful under NETA because it can be used to change the load shape, reducing or shifting customer's demand away from the periods of volatile prices or demand. This could potentially reduce the level of price and volumetric risk. Load shape modifications are achieved using one or a combination of the following methods illustrated below in figure 4.11⁵⁰. The x-axis is time and the y-axis is load.

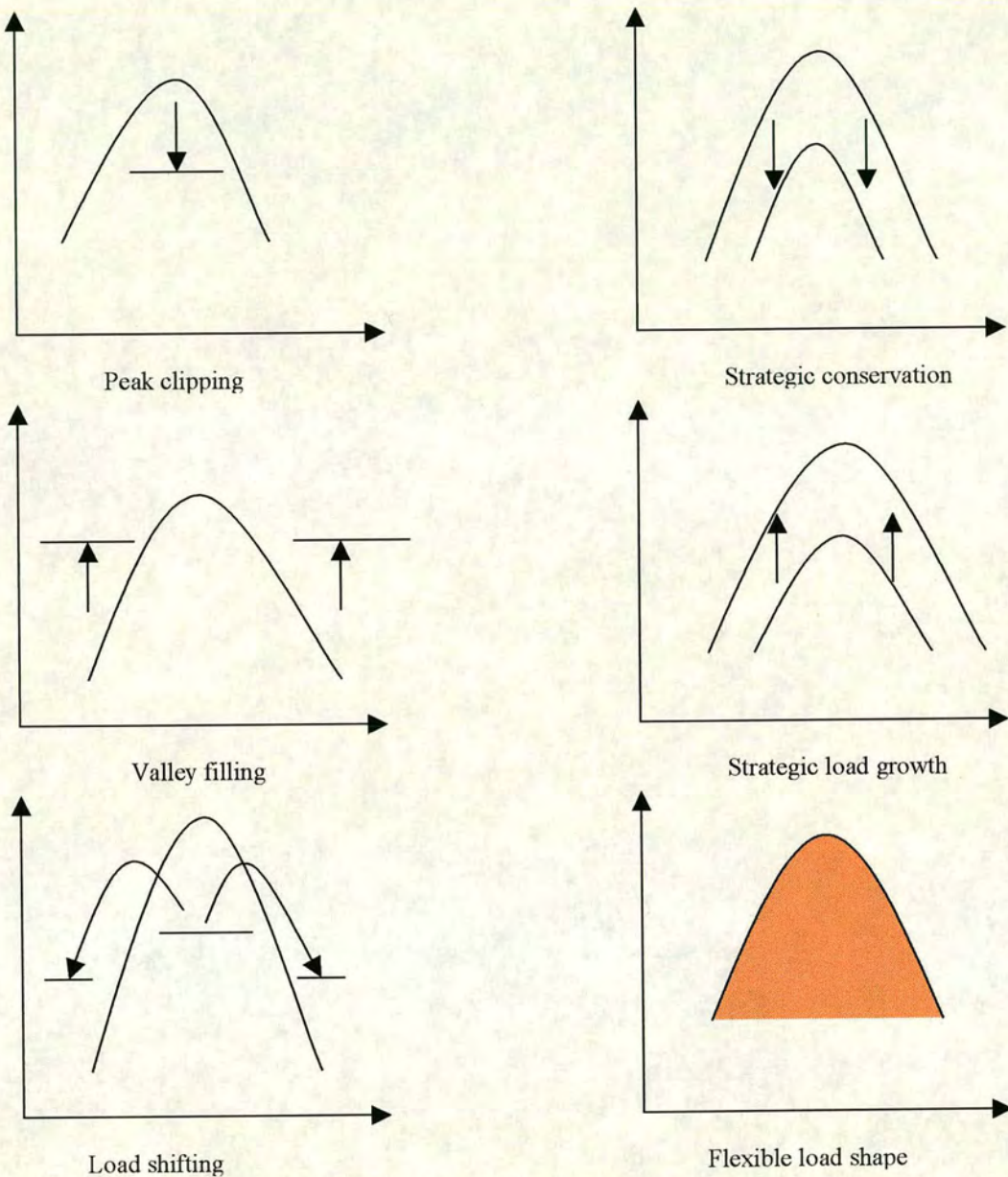


Figure 4.11 load shape modifications with the x-axis referring to time and y-axis to load

- **Peak clipping** - A reduction in electricity use usually restricted to times of near system peak conditions
- **Valley filling** - An increase in off-peak use
- **Load shifting** - A shift in use from peak to off-peak periods
- **Strategic conservation** - A reduction in overall use of electricity
- **Strategic load growth** - An increase in overall use of electricity
- **Flexible reliability** - An implied reduction in the availability of energy reflected in an increased probability of service being interrupted.

Suppliers can achieve load shape changes through a variety of methods. The different types of incentives available to Suppliers are outlined below:

- **Load control:** this involves controlling customer energy usage. This is done either through direct, local, or distributed methods, through the installation of communication devices at both Utility and customer facilities, enabling real-time alteration of the load
- **Time-of use rates:** Higher rate tariffs during peak demand hours can encourage customers to conserve and to shift demand to off-peak hours.
- **Interruptible and Curtailable rates:** In exchange for discounted rates, customers with typically large industrial or commercial facilities allow power to be turned down or off when it is needed elsewhere in the system
- **Thermal energy storage (TES):** These systems use off-peak electricity to heat a storage medium, and then draw from that medium to provide space heating during peak hours. TES systems offer customers energy savings where time-of use rates or demand charges are in effect.
- **Cogeneration:** When capacity is constrained, some utilities promote customer use of on-site production of electricity. Though this might appear to reduce host utility revenue, many of these units are owned and leased by the utility. An example of Cogeneration is Combined Heat and Power (CHP) where there is simultaneous generation of usable heat and electricity in a single process. The heat generated in the process is utilised via suitable heat recovery equipment for a variety of purposes including: industrial processes, community heating and space heating⁵¹.
- **Conservation:** Energy conservation programmes encourage customers to improve the thermal integrity of buildings and to use more efficient equipment and appliances to save energy costs. This results in decreased demand. To achieve this, some utilities give discounts as an incentive on energy conservation devices.

So far there has been fairly limited use of DSM under privatisation except in Scotland. This has been for reasons given earlier. Under NETA, there will be a greater incentive for Suppliers to manage their load more effectively, especially for smaller customers because of the volume risk they pose. This could lead to more complex and dynamic tariffs and also more creative use of DSM, to reflect more accurately the time of day costs and to control possibly demand⁵².

4.8.6 Portfolio management

In the financial markets, if the pattern of return on two risky stocks is sufficiently uncorrelated, then combining the returns on both stocks will reduce the risk, as the fluctuations cancel each other out. The percentage of risk on individual shares decreases by combining stocks with uncorrelated returns cancelling out the firm specific risk. The risk that remains within the portfolio is the market wide risk and is common to all shares. The optimal portfolio size is 15 stocks, beyond which the risk reduction is negligible⁵³.

This technique of combining assets to reduce overall risk has many applications. By gathering a large diverse customer base, Suppliers eliminate some of their volumetric risk. Random demand variations from individual customers cancel out each other leaving market wide effects like the weather and price risk. This technique goes some way towards reducing an element of volumetric risk and would have even greater relevance under NETA by reducing some of the volumetric risk.

4.8.6.1 Value-at-risk

The concept of Value-at-risk (VAR) was developed as a response to concerns about high profile derivative losses in the financial industry in the early 1990's⁵⁴. J.P Morgan introduced the concept in December 1994 as part of a package known as RiskMetrics⁵⁵. VAR is the amount of money a trading company might lose due to the market risk of asset prices moving against it before it can close its position. In practical terms, VAR measures, for a portfolio, the worst expected loss to a specified confidence level during a given period of time under normal market conditions. VAR effectively gives a measure of the level of a company's market risk exposure. For its calculation, the concept uses

- The sensitivity of a portfolio to changes in underlying prices, which reflect how well the portfolio is hedged (the more fully hedged the less sensitive it is to price changes)
- The volatility of prices, which reflects the likelihood of large price changes.

VAR can be applied to many risk management related activities with varying objectives. The main criticism of this method is that it does not take into account abnormal events like market crashes. In the US, VAR has been applied to the electricity market with a varying degree of success. It must be noted that there are fundamental differences between the electricity and the financial markets which complicate the application of VAR to the electricity markets⁵⁶. For example, the electricity spot price follows a mean-reverting process and has very strong annual and semi-annual seasonality factors. Added to this, VAR calculations only account for price risk and using VAR on real energy portfolios does not account for volumetric risk. Care should be taken when attempting to apply this method to NETA, because it requires a large historic data set to infer robust volatility and correlation calculations.

4.8.6.2 Profit-at-risk

Most electricity Suppliers focus on sales margins from their retailing activities as the key performance measure for management reporting. These are profit or revenue figures, which are very different from those of a bank, which focuses on portfolio evaluation, and the daily management of the balance sheet. VAR for a bank gives a risk measure that is directly related to the balance sheet value. As a Suppliers core business is measured in terms of profits and margins, a risk measure that is directly related to profits is more appropriate. This has led to the emergence of 'Profit-at-risk' (PAR) as the key measure of risk in the energy industry⁵⁷.

The main feature of PAR is that, unlike VAR, it assumes that positions will be taken through to delivery rather than be closed out. PAR works by using simulation-based modelling, which basically tests the whole range of risks affecting spot prices at the time of delivery. It models volume risk by running many different simulations, which allow parameters to be set for safe trading.

The above two portfolio management methods, VAR and PAR are very recent additions to the risk management arsenal and access to knowledge about their use within the UK's ESI is scant because of commercial sensitivity.

4.8.7 Metering

Half-hourly metering provides Suppliers with an effective way of determining customers historic demand patterns. Half-hourly metering will be very necessary under NETA because it is a means by which a Supplier can determine a customer's volumetric risk. This enables not only Suppliers to settle the cost of the contract more effectively by passing on the cost of demand variation to the customer, but to also monitor the performance of the contract during its time of operation, through remote metering⁵⁸. Though metering will not help the performance of the contract, since generally tariffs have been agreed beforehand, the Supplier will be alerted to deviations from expected behaviour and can take steps to limit the risk through, for example, buying extra EFA cover.

Half-hourly metering provides Suppliers with useful information but the cost of meter installation and the administrative resources required to manage and maintain data collection and billing, make effective meter provision to all customers impractical. This is particularly true for smaller customers where the price of effective half-hourly metering makes up a significant proportion of their electricity costs. Ironically, this is the group that provides the most variation. Instead, Suppliers use standard profiles to calculate tariffs for each class of low demand customer⁵⁹. Chunking and Algorithmic Profiling are performed to make the tariffs more reflective of real time

costs. There are several innovative metering options coming on the market place in anticipation of NETA.

4.9 The future of risk management under NETA

This chapter has concentrated on describing the risks Suppliers face within the ESI, the methods used to quantify and manage these risks. NETA will present increased risks to Suppliers. The industry will have to evolve a new generation of more sophisticated risk management tools and methods. This section discusses the future of risk management in the ESI and recommends suitable strategies.

4.9.1 Price risk management

In the present system, Pool price has been the main risk which majority of the present generation of tools have evolved to manage. These tools and methods would be inadequate for NETA for the reasons discussed in this section.

NETA will be based on a series of bilateral markets. They will be starting from scratch and will have no semblance to the Pool mechanism. Price determination might be done separately in the various markets and could be influenced by a wider range factors than in the Pool. For example, in the case of extreme weather forecast, the 24-hour prices in bilateral market could be rather high. Closer to real-time most of the demand would have been met through the 24-hour market. Balancing Market prices may not reflect the 24-hour market's price increases. Currently, it would be impossible to forecast the prices in the various markets until adequate data has been collected. A large database, robust enough to perform accurate price forecasts on which exposure calculations are based will take a long time to gather.

Most CFD and EFA contracts were referenced to one or a combination of the components of Pool price. NETA will have no single reference price; prices will be determined bilaterally in the various markets. Contracts might be priced based on a combined basket index of the spread of prices in each market or made up as a weighted combination of average prices in the various markets. Either way, the future pricing and referencing methods for long-term bilateral contracts is not clear. Any new methodology could have a new set of hidden risks

Portfolio based techniques like VAR and PAR will be fairly redundant because they rely on historic data to determine the correlations between the assets or the contract values within the portfolio. These will not be available until several years after NETA's inception. For these reasons, it will be very difficult to design a robust strategy for price risk without making sweeping generalisations which could be inherently flawed.

4.9.2 Volumetric risk management

The causes and components of volumetric risk will stay the same regardless of the market mechanism or price-setting model and so too would the risk management tools, like weather derivatives or demand driven swing contracts. However, the financial costs of volumetric risk do change according to the market mechanism. From the discussion in the previous chapter, one of the basic tenets of NETA is that Suppliers should be more accountable to demand shortfalls through balancing market prices and steep imbalance penalties. This means that the implications of volumetric risk will be much higher than in the Pool.

Suppliers will be required to do real-time balancing, and will face penalties for getting it wrong. In the Pool Suppliers had limited knowledge of demand behaviour because their primary function was to sell electricity and to bill customers. Balancing was done solely by the system operator and Suppliers were billed for any demand shortfalls. In NETA, Suppliers will be required to submit their Final Physical Notification (FPN) to the system operator by gate closure, which is 4 hours before the half-hour in question. The FPN's should provide information on the exact requirements at BSP level. Suppliers will have to gain intimate knowledge of their customer behaviour to provide this information. In the Pool demand forecasting was performed primarily by the NGC at a national level and as such, it gave no information on demand requirements at BSP level. BSP forecasting has not been performed actively in the UK.

Accurate demand forecasts would effectively be used by Suppliers to quantify their volumetric risk from which the exposure to imbalance prices will be calculated by subtracting the forecasts from the existing contract position for that half-hour in question. From the exposure calculation, Suppliers can make up any demand shortfalls either through nominating demand-driven swing options or in the balancing market. Accurate demand forecasting at BSP level will be invaluable to Suppliers in NETA.

The remainder of this thesis describes work carried out to develop an optimal BSP forecasting model over a 4-hour time horizon to provide the SO with accurate demand expectations at Gate Closure. The Supplier would use the forecast to make up the difference between the contracted volume and expected demand to avoid imbalance charges. In the subsequent chapters, the various methods available for forecasting demand are compared, discussed and the most appropriate ones analysed. From this, the chosen method is optimised with a view to improving Suppliers volumetric risk measurement.

4.10 Summary

This chapter introduced the concept of risk and its measurement and management. Suppliers faced a variety of risks from their trading activities including price, basis, credit, operational and volume risks during the Pool's existence. These are expected to increase under NETA. The most commonly used methods for measuring price and volume risk, the greatest risk Suppliers faced under the Pool was outlined. To introduce the concept of risk and its management in the ESI, the range of tools and methods evolved since privatisation to manage price and volumetric risk were presented and their use summarised. Possible modifications of the existing risk management tool were suggested for NETA

The arrival of NETA will invalidate the majority of these tools in their present incarnation. Modifications to existing methods, together with a new generation of tools, will be needed for effective risk management in an uncertain era. Designing tools for price risk management would be pointless because at the present time there is no historical data or prior knowledge of the markets operation. Volumetric risk stays the same regardless of the market configuration. Using optimal short-term demand forecasts, combined with knowledge of contractual positions, will alert Suppliers to exposures, allowing them to be managed. In order to avoid NETA's imbalance charges, accurate demand forecasts are necessary for the elimination of volumetric risk.

5. Electricity demand forecasting

5.1 Overview

Electricity demand forecasts are essential for the effective operation of the ESI. The system operator, National Grid Company (NGC), traditionally performed this function. These forecasts were used for scheduling, maintenance and long-term system planning. With the introduction of the New Electricity Trading Arrangements (NETA), Suppliers are required to submit Final Physical Notifications (FPN) 4 hours in advance to the system operator of their half-hourly demand position at every Bulk Supply Point (BSP) they service. Failure to provide accurate demand figures results in heavy penalties, which could have severe financial implications to Suppliers. As a result, accurate demand forecasting will become a vital part of Suppliers risk management activities.

This chapter examines demand forecasting with a view to devising the appropriate forecasting tool for Suppliers needs. Before accurate demand forecasts can be made, the demand has to be understood together with the factors influencing its behaviour. The different types of demand forecasting methods are discussed in the second half of this chapter to determine which ones can be applied to satisfy NETA's requirements.

5.2 Customer demand

In December 1933, George J. Read of Brooklyn Edison Company wrote in *Electrical World* magazine "urgent need for intelligent forecast of the future electrical demands and consumption is unfortunately accompanied by factors which weaken faith in most of the various methods used for past forecasts"⁵⁹. This statement remains true today and serves as a reminder of the problems and difficulties of predicting the future. Human needs ultimately drive demand for energy, but the great variety of its consumers and end users creates a wide range of dissimilar use. Figure 5.1⁶⁰ shows the major customer segments for electricity demand in the UK: domestic, industrial and commercial. Each of these demand sectors will be discussed in more detail in the following sections.

5.2.1 Industrial users

Industrial users constitute up to 28% of the total national demand with the majority of this load centred in areas with high industrial activity. The bulk of this load is used up in industrial processes, for example furnaces. The industrial load demand curve is very predictable and tends

to follow similar daily patterns. Excluding anomalies, this demand segment is not very variable and does not present to the Supplier much uncertainty in demand.

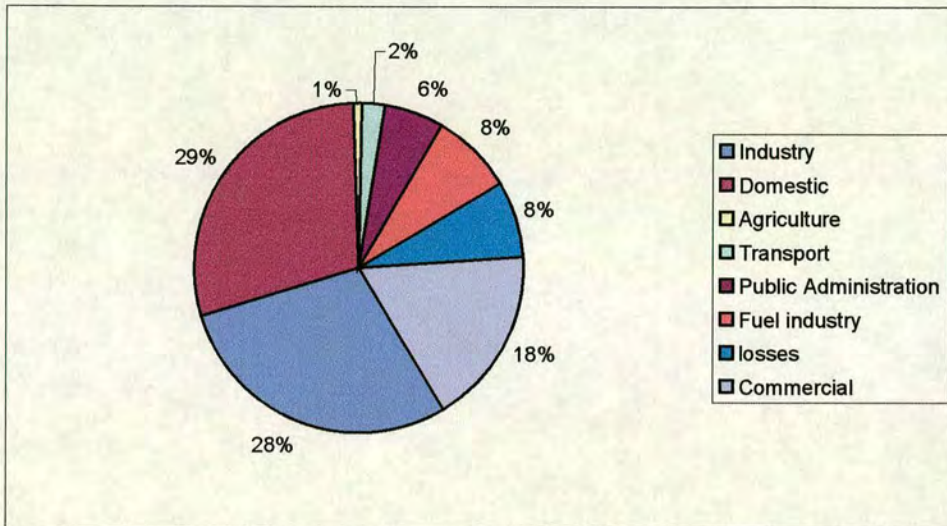


Figure 5.1. Electricity demand by sector 1999 total demand 43.32GW

5.2.2 Domestic/Commercial users

This segment makes up 47% of the total demand. Commercial customers who include hospitals, banks, shops etc have been included with domestic users because a high percentage of their demand is influenced by the same fundamental factors, notably the weather. Customer consumption patterns in the domestic segment are very variable. On any given day no two households will have identical demand, and even within a given household consumption patterns vary from day-to-day. The domestic customer segment causes the most demand uncertainty, hence volumetric risk to Suppliers.

In addition, customers in both segments are not half-hourly metered. Thus their demand uncertainty cannot be apportioned and recovered. They will cause the greatest level of volumetric risk to Suppliers, as such it was decided to focus on a Bulk Supply Point (BSP) with a predominantly domestic/ commercial load for the analysis to select the best forecasting method.

5.3 Fundamental demand drivers

Electricity demand is controlled by the dynamic effect of its fundamental drivers. They are the factors that cause consumers to increase or decrease their consumption. Understanding these factors is essential to forecasting demand.

5.3.1 Economic factors

The economic climate in which a Supplier operates has an effect on electricity demand patterns, with the rate of demand growth or decline heavily dependent on economic trends. Increased economic activity also leads to higher investment in domestic, industrial and commercial activities. The reverse occurs during economic downturns.

Total domestic demand increased by 25 per cent between 1970 and 1998⁶², due largely to an increase in the number of homes which in itself is reflective of the increased economic activity over the period. Though there has been an increase in the number of electricity appliances during this time, this was offset by greater efficiency. Commercial demand also increased by 17 per cent during this period, mainly due to the ever-increasing reliance on electrical equipment and the growing use of air conditioning.

Because they operate over long time horizons and do not have a direct effect on a daily basis, economic factors are not explicitly represented within short-term forecasts. It is very important to recalibrate short-term forecasting models frequently to capture the underlying trends.

5.3.2 Time factors

Electricity demand is very dependent on the time of use. The four principal time related factors that play an important role in influencing load patterns are the seasonal effects, day of the week, time of day and special days. These will be discussed in greater detail in the following sub-sections.

5.3.2.1 Seasonality

Demand changes significantly over the year. This is illustrated in figure 5.2 which shows the average daily demand in 1998 of a predominantly domestic BSP in north eastern Scotland.

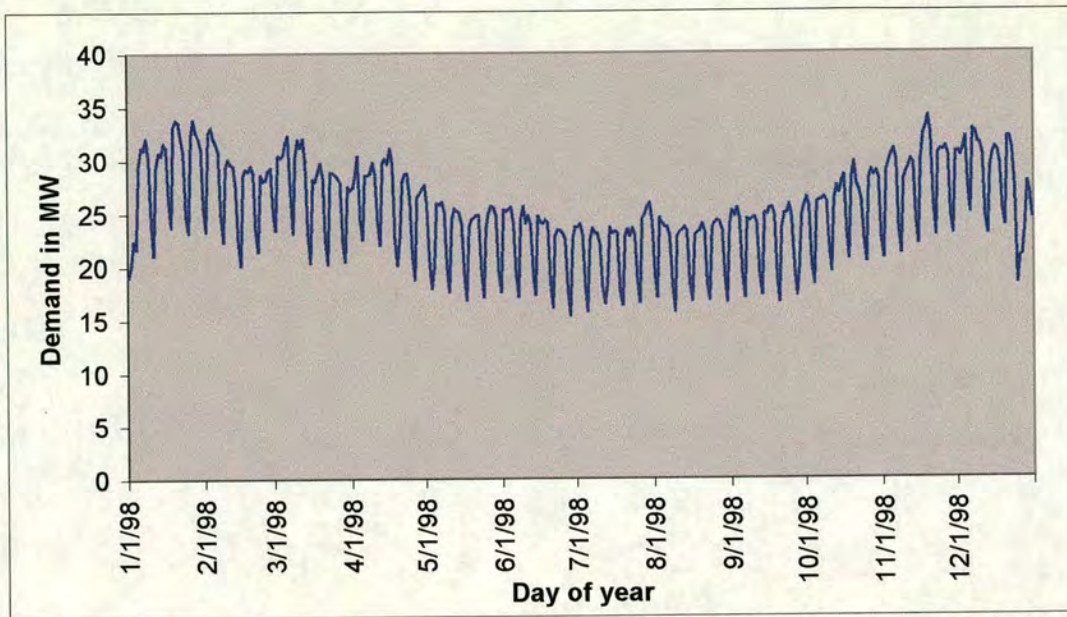


Figure 5.2. Average daily demand for 1998

The load is cyclical with the peaks on the graph representing the weekdays and the troughs, the weekends. Over the year, the demand pattern shows a cyclical pattern with higher average daily demand during the winter, gradually descending till the beginning of July, then ascending again towards the end of the year. The yearly load cycle occurs gradually in response to the length of daylight hours and average temperature changes. The length of daylight hours controls the magnitude and duration of lighting peak. Figure 5.3 illustrates the average domestic household lighting demand by season. The winter lighting demand peak is very high and occurs in the late afternoon, coinciding with domestic customers arrival home from work⁶³. As the length of daylight hours increases, the magnitude of the evening lighting peak decreases and occurs progressively later in the evening. The timing of the morning lighting peak is constant throughout the year but its magnitude reduces to almost zero in mid summer.

The daily average temperature affects both space and water heating loads. It follows a periodic cycle over the course of the year. In countries with a high air-conditioning load, for example USA, demand shows a significant summer increase. The summer load is still insignificant in the UK, but there have been signs in recent years that things could be changing because of the increased use of air-conditioning and large computer loads.

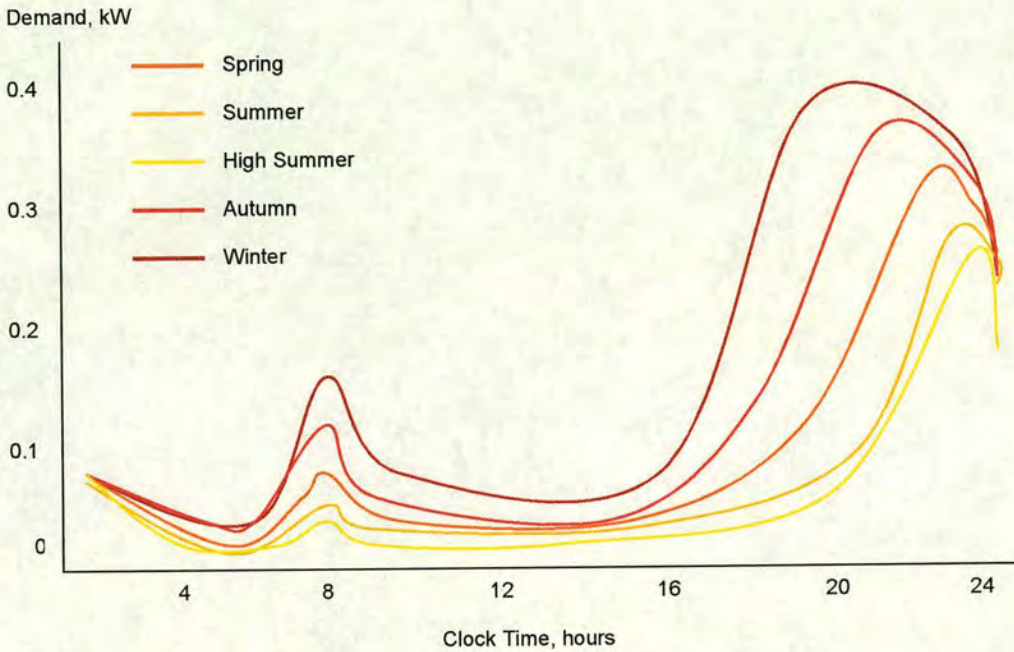


Figure 5.3. Weekday lighting demand by seasons

5.3.2.2 Time-of-use

Domestic and Commercial demand varies considerably across the day, either as a direct or indirect result of daily work patterns. Daily demand tends to be fairly flat except during the two daily peaks. The morning peak corresponds to when domestic customers are preparing for work and the evening peak commences when customers return from work. The evening peak lasts till around 8:30 pm, after which point demand begins to tail off. There is a demand increase around midday, corresponding to the lunch break.

In some parts of the country Demand-Side Management (DSM) has been used to alter the daily demand pattern described above. For example, in northern Scotland DSM is used to distribute blocks of the space-heating load across the day. This was done in an attempt to flatten the load curve to optimise generation. As a result the morning and evening peaks are less noticeable because demand is more evenly spread over the day.

There is considerable demand variation across the days of the week. The demand patterns on weekdays tend to show similar trends but vary from the weekend curve. The average magnitude of the weekend curve, particularly Sunday, is lower because majority of the commercial load is absent. As a result, consumer demand is more evenly spread across the day resulting in less pronounced evening and morning peaks. The diagram in figure 5.4 illustrates the day of week demand at the BSP in northeastern Scotland. The curve shows a high demand during working

hours, corresponding to a high commercial load and shifted domestic heating load as mentioned earlier.

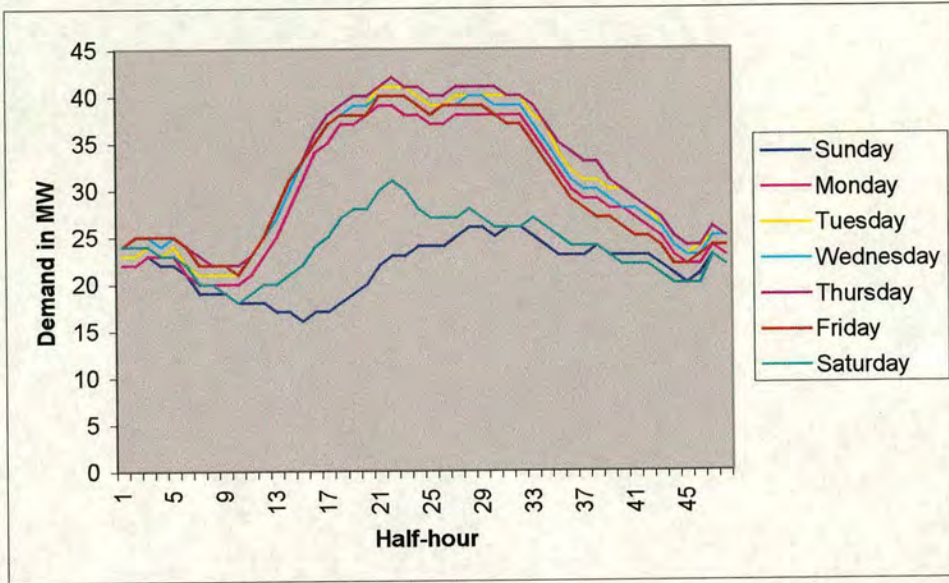


Figure 5.4. Daily demand for week starting 4th January 1998

5.3.2.3 Special days

On special days, for example public or religious holidays, the demand pattern is significantly different, due to the absence of commercial demand and altered domestic customer use. The average demand is lower and tends to have a similar shape to Sunday's curve (refer to the figure above). Special days must be taken into account in short-term demand forecasts.

5.3.2 Random factors

The system load is continuously subject to random disturbances reflecting the fact that it is made up of a large number of diverse individual demands. For this reason, the load will always have an inherent degree of randomness. In addition to these, there are the unpredictable one-off events that can greatly alter the load. These include industrial action, notably strikes or shut downs and special TV programmes. There was a noticeable demand spike during the half-time break of Scotland's Euro 1996 game with England⁶⁴. This was not caused by the power consumption of TV's, which is comparatively very low, but due to kettle consumption during the half-time interval. Another example of a random demand change was the Solar eclipse in August 1999. The national demand pattern on the day, from 9am to 1pm are shown in figure 5.5, these are contrasted with the demand pattern for the same period on the previous day. The demand shows a dip between 11:00 and 11:45am, the period of the total eclipse. This was because majority of the industrial and commercial load was shut down as the general public took a break to view the

eclipse. The effect of these unpredictable events on the load magnitude is very uncertain and is impossible to forecast. However, because large demand altering events occur very infrequently they do not present an everyday risk and as such not modelled within the forecast.

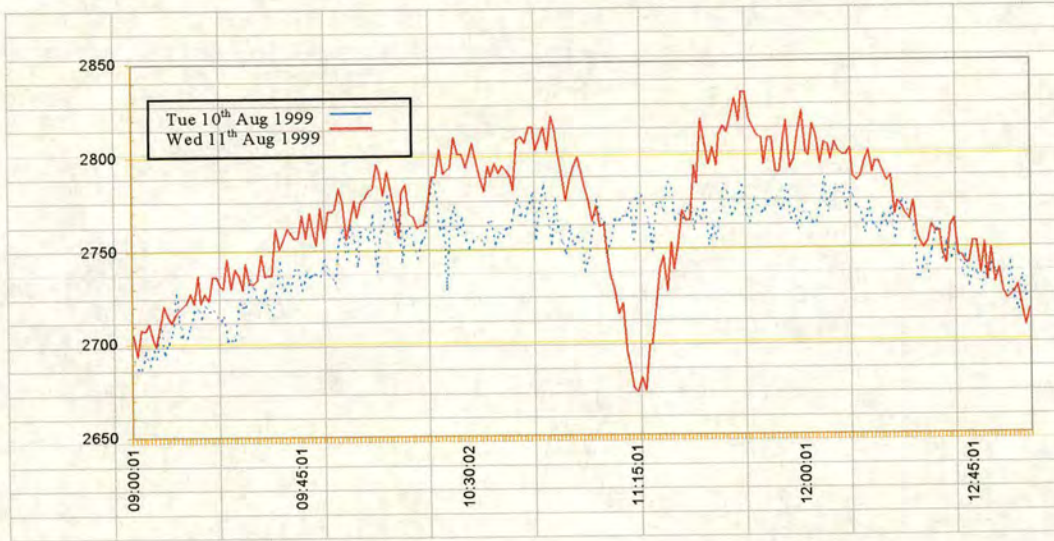


Figure 5.5. Demand during solar eclipse

5.3.3 Weather effects

The weather is the largest fundamental driver of domestic and commercial demand. The weather factors that have been found to influence the demand are as follows⁶⁵:

- Temperature: Directly affects the level of space and water heating demand.
- Wind speed: Affects the space and water heating demand. Wind speed is responsible for rate of heat dissipation from the external surfaces of buildings. When the outside temperature is low the wind speed increases the rate of conduction. At high outside temperatures, conduction is lower so wind speed has less effect.
- Illumination effects: This is related to the level of cloud cover and the length of daylight hours. The illumination effect determines the lighting demand
- Precipitation: This factor also directly affects the rate of heat loss from building surfaces. In cold weather, the water effectively serves to help cool down buildings. Precipitation has an indirect effect of increasing domestic consumption by keeping people indoors where they will use more electricity.

Temperature is the weather variable with the greatest effect on demand. During the winter, a fall in temperature of one degree Celsius resulted in an increase of about 3% in average load during the half-hour 17:00-17:30 for British unrestricted domestic consumers in the year 1990-91⁶⁶. Figure 5.6 shows a scatter plot of the demand against temperature at the BSP in northeastern Scotland. The demand-temperature relationship may be seen.

On the scatter plot, there are two similarly shaped clusters: the denser cluster represents the weekday demand, which is of a higher magnitude on average because it includes the commercial demand. The anomalous point on the scatter plot with zero demand at 9 degrees was caused by maintenance work on the BSP.

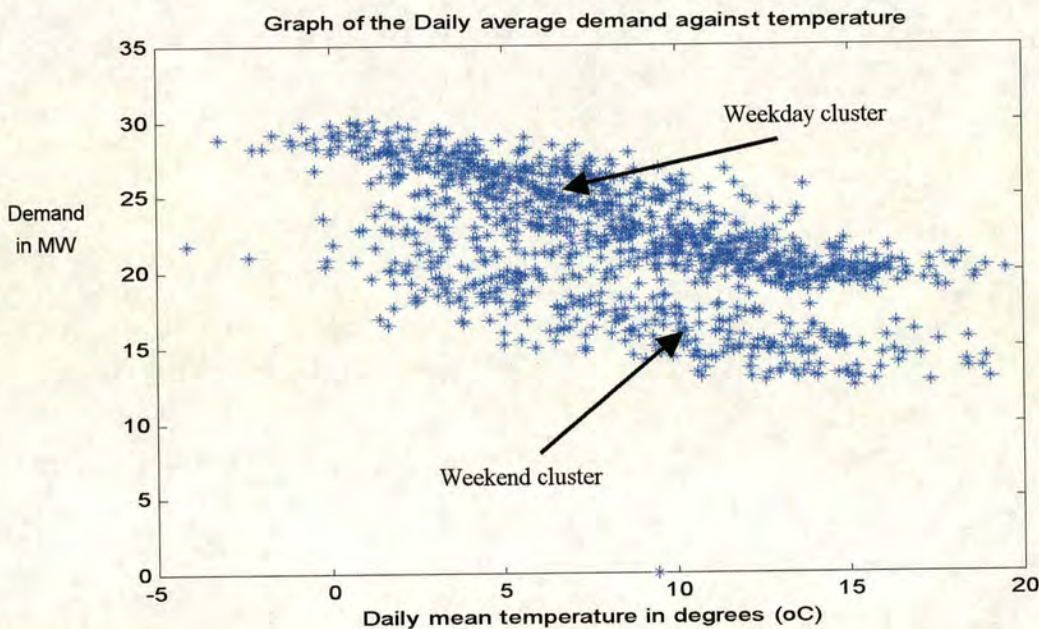


Figure 5.6. Daily average demand against temperature

Electricity is very widely used in the domestic and commercial environment. The next section examines the weather sensitive components of domestic and commercial demand.

5.3.3.1 Space heating

Space heating contributes up to 75% of the total domestic electricity consumption. The space heating demand within a Supplier's BSP region usually depends on the local penetration of gas as a heating fuel. In northern Scotland, electricity is the principal fuel for heating purposes because the gas network is very limited.

The relationship between the space heating energy demand and outside air temperature depends on the following: a physical loss effect, socio economic factors and other individual consumer choices notably, the desired level of thermal comfort⁶⁷. This section briefly explains this relationship.

Three forms of cooling determine the rate of heat loss from the dwelling space: conduction, convection and radiation. The heat loss Q from the outside surfaces can be viewed as an increasing convex function of the difference between internal temperature, T , and the outside temperature, T_i , and can be represented by a polynomial⁶⁸

$$Q = c_1(T - T_i) + c_2(T - T_i)^2 + c_3(T - T_i)^3 + c_4(T - T_i)^4 \quad \text{Equation 5.1}$$

where c_1 to c_4 are coefficients.

The convexity of the heat loss function means that, as outside temperature increases less energy is required to maintain the desired level of thermal comfort. Thus, the relationship between the demand and outside temperature becomes less marked as the temperature approaches this level. The level of thermal comfort depends on the consumer's preference and ability to afford it. For example, a retired single pensioner might want to keep warm but would opt for a lower level of thermal comfort to avoid high space heating cost.

Thus, the implicit price of thermal comfort falls as outside temperature increases because consumers elect to switch off their heating systems. There is a temperature above which temperature has no effect on demand. Some consumers may change their behaviour in response to lower temperatures by wearing warmer clothes indoors. The net outcome of the price effect and the consumer's level of thermal comfort on the space heating demand will mean that as temperature decreases the increasing gradient of the convex relationship between temperature and demand will be dampened. These effects will vary according to the individual households.

Another determinant of the relationship between the demand for space heating and outside temperature is the capacity of the heating system. If at low temperatures, the heating system saturates and cannot maintain the difference between the customers desired temperature and outside temperature, the relationship between demand and temperature will flatten.

The rate of heat loss from buildings is affected by the following factors.

- **Quality of insulation:** This is dependent on the materials used in its construction and the area of the exposed surfaces. Older buildings tend to be less well insulated and have larger external surfaces and as such, they lose more heat.
- **Duration of the cold/hot spell:** The temperature has a lagged effect on the rate of heat loss because it takes a period of time for building surfaces to adjust to the external

temperature. The longer a cold spell lasts, the greater the heat loss. This is the reverse during warmer spells. There is an exponential relationship between the duration of the cold/hot spell and the rate of heat loss/gain⁶⁹.

- Wind speed: As mentioned earlier, wind speed controls the speed of the external airflows. Heat loss increases with higher wind speeds.
- Rate of heat build-up/dissipation: This is related to the building density. Densely built-up areas have self-insulating properties because their radiated heat warms up the surrounding air, slowing down the rate of heat loss. Built-up areas also act as windshields, reducing the rate of heat conduction from building surfaces, increasing the rate of heat build-up. Conversely, isolated dwellings have a higher heat-dissipation rate. Location and building density have an effect on heat loss thus heating demand.

The discussion above suggests that there are various complex physical, economic and consumer preference factors that influence the relationship between the space heating demand and outside temperature. These effects are non-linear, complex and thus difficult to model.

Space heating demand is very sensitive to the time of the day. In the UK there are two domestic tariffs: either a time-invariant tariff or a tariff with a reduced over night price but a higher standing charge (Economy 7). The majority of Economy 7 customers are supplied with their heating demand at night to take advantage of the cheaper rates. This increases the overnight weather- demand sensitivity.

The time-invariant customer would have the preference to use space heating at any given time. Commercial customers generally tend to use majority of their space heating during business hours with the notable exception of hotels and hospitals. The space heating demand sensitivity to the weather will be higher during the night because majority of domestic customers are indoors.

5.3.3.2 Water heating

The water-heating load is heavily weather dependent. It is influenced by both physical losses and consumer thermal preferences. In water heating, physical losses are primarily from the hot water cylinder, the primary pipework and the distribution pipework. The losses from the hot water cylinder depends on

- The type and thickness of the cylinder
- The temperature of the stored water
- The volume and shape of the cylinder
- The length of time before hot water is drawn off.

Heat losses from the primary pipework can be significant, particularly if the pipes are uninsulated or the system keeps the pipes hot for long periods of time. Distribution losses between the cylinder and the tap are fairly large and are about 15% of the energy leaving the tank⁷⁰. The weather-water heating demand relationship is very sensitive to the time-of-day during weekdays because the heating load coincides with the daily demand peaks. The relationship is less sensitive to the time of day at the weekends as water-heating demand is more evenly spread across the day. Water heating is a less significant load than space heating and makes up less than 20% of the total heating requirements.

A percentage of the heat lost from water heating contributes towards space heating. However, dwellings with poor space heating will have higher water heating losses on average.

5.3.3.3 Refrigeration and freezers

The electricity consumption of fridges and freezers is related to the surrounding temperature. They cool their contents using the principle of latent heat of vaporisation. As the surrounding temperature increases, they would require more energy to maintain the lower internal temperature. The relationship between fridge /freezer energy consumption and temperature is non-linear⁷¹.

5.3.3.4 Others

There are various other appliances within the domestic household that are affected either directly or indirectly by the weather. General consumption goes up during the winter and over bad weather spells because consumers stay indoors longer, increasing their overall usage. For example, more kettles are boiled during cold, wet periods than at the height of summer.

5.4 Forecasting

The fundamental challenge to forecasters was best described by the Nobel-laureate physicist Neils Bohr when he stated, “ prediction is a very difficult art especially when it involves the future. Like all statistical methods, demand forecasting is both a science as well as an art: The science and art of specification, estimation, and evaluation of the models, which depends both on the model’s effectiveness and the operator’s ability”.

In this discussion only the model’s effectiveness at forecasting will be considered. The question of operator competence will be assumed to be the same with all models.

Modern demand forecasting uses a variety of statistical methods to predict electricity demand. These methods include econometrics, time-series and artificial intelligence techniques. This

section presents an overview and discussion of these methods. Firstly the forecasting time frame is considered.

5.4.1 Forecasting time-frame

Prior to NETA's inception demand forecasts were performed primarily by NGC. Three types of forecasts were performed over short, medium and long term time horizons. Each type was used for different purposes and was influenced by different factors which required different techniques. The three forecasting time horizon and their uses are discussed in turn:

- **Short-term forecast:** These were used by NGC under the previous system primarily to provide dispatcher information, which was used for generator scheduling and to perform off-line security analysis. The timescale for short-term forecasts vary from a few minutes to one week. In the short-term, demand is primarily influenced by the weather, day of week, time of day and random events, notably television schedules. Forecasts are usually made either every hour or half-hour. Economic and seasonal factors do not directly affect short-term demand because they work over longer time horizons.
- **Medium term forecast:** These were used to provide information for risk management purposes, for example Generators priced and sold Contracts For Difference (CFD) based on future demand expectations. Medium term forecast were also useful for scheduling generator and grid maintenance or repair⁷². The timescales for medium-term forecasts vary from one week to a year. The principal demand drivers over this time horizon are day of the week, weather, seasonality, and economic factors. Average weekly and monthly weather variables have a greater effect in the medium-term than daily weather changes. The forecasts are made either daily or weekly and is not done half-hourly because the variables determining half-hourly demand are not known with much certainty over time horizons longer than a week.
- **Long-term forecast:** They are used to plan long-term government energy policy and to enable market participants to make long-term investments. Long-term forecasts are used by Suppliers to target demand growth sectors. The timescales for long-term forecast range from a year to any specified time in the future. The principal demand drivers are economic factors and seasonality. Forecasts are made either weekly, monthly or yearly.

For the purposes of this discussion, short-term forecasting only will be considered because under NETA, Suppliers would be required to provide demand predictions at Gate Closure (GC), 4-hours in advance of the scheduled half-hour.

5.5 Forecasting methods

This section examines the wide array of forecasting methodologies available from the four most common generic classes; Regression, Time-series, End-User and Artificial Intelligence. This is done to assess the merits and demerits of each method in meeting Suppliers NETA forecasting requirements.

Each of these methods has been used extensively in academic literature. However, as at the time of writing a comprehensive review of their implementation and application to NETA did not exist. This section attempts to achieve this by describing and assessing the various methods from the large volume of work available. The two most appropriate methods to meet NETA's requirements are selected and compared analytically in future chapters.

5.5.1 Regression methods

Regression models are used to find the relationship between independent variables and a corresponding dependent variable. In the case of the ESI the independent variables comprise of weather, time and previous demand and the dependent variable is future demand. Regression methods are grouped into two distinct classes: Parametric and Nonparametric.

Parametric regression methods formulate a mathematical or statistical model by examining qualitative relationships between the dependent variable, load, and load affecting factors. The model's parameters are estimated from historical data and the adequacy of the models is verified by analysis of the model's residuals, i.e., forecast errors. Parametric models include Multiple Linear Regression (MLR) and Polynomial Regressions. With nonparametric methods the relationship is determined graphically, directly from the historical data set. Kernel smoothing is the most popular nonparametric method for short-term forecasting.

5.5.1.1 Multiple Linear Regression (MLR)

This is the most commonly used short-term forecasting method in practice because of its flexibility and ease of application. MLR is either used as a stand-alone method for load forecasting or to define the transfer function in a time-series load-forecasting model. The National Grid Company uses MLR to perform short-term national demand forecasts. With this method, the load forecast is determined in terms of explanatory variables such as the weather and non-weather variables. The MLR will take the form of:

$$y(t) = a_0 + a_1x_1(t) + \dots + a_nx_n(t) + a(t) \quad \text{Equation 5.2}$$

where

- $y(t)$ = Electrical demand at time t
 $x_1(t), \dots, x_n(t)$ = Explanatory variables correlated with demand
 $a(t)$ = A random error term with zero mean and constant variance
 a_0, a_1, \dots, a_n = Regression coefficients

Correlating each of these variables with the load variable identifies the explanatory variables to be used in this model. The estimation of the regression coefficients is usually found using either least squares estimation or maximum likelihood estimation. Statistical tests such as the F-statistic and t-ratios are performed to determine the significance of the regression coefficients.

As the name implies, MLR gives a straight-line approximation of the relationship between the demand and the explanatory variables. It assumes that the relationship is linear. From the discussion in the previous section, the demand-weather relationship is not linear. Thus this method might not adequately illustrate the relationship and could bias the forecast.

There are various methods to adapt MLR to non-linear relationships between the dependent and explanatory variables, notably, transformations of linear into non-linear term using an appropriate arithmetic function that adequately expresses the relationship. Below are examples of commonly used mathematical transformations. The two-variable regression model is used in the illustrations.

- log-linear: this involves a logarithmic transformation of both the dependent and explanatory variable.

$$\ln y(t) = a_0 + a_1 \ln x_1(t) + a(t) \quad \text{Equation 5.3}$$

- Semilog: The explanatory variable is in linear form and the dependent variable is in log form.

$$\ln y(t) = a_0 + a_1 x_1(t) + a(t) \quad \text{Equation 5.4}$$

- Lin-log: the dependent variable is in log form and the independent variable in linear form

$$y(t) = a_0 + a_1 \ln x_1(t) + a(t) \quad \text{Equation 5.5}$$

- Reciprocal: This model is appropriate when the relationship between the explanatory and independent variable can be explained by reciprocity.

$$y(t) = a_0 + a_1 x_1(t)^{-1} + a(t) \quad \text{Equation 5.6}$$

To determine whether or not to use a mathematical transformation and if so, the most appropriate transformation to use, a scatter plot should be plotted for the dependent variable $y(t)$ against the independent variable $x(t)$. A visual examination of the scatter diagram is made to select the most appropriate transformation.

There are several considerations to be made regarding the transformation of $x(t)$, most importantly whether a transformation actually results in a significant improvement of the model. When there is more than one explanatory variable, a multidimensional scatter plot is required. If there is correlation between the explanatory variables, care must be taken when interpreting the scatter diagram because it is very difficult to determine the exact contribution of each variable. This could obscure selection of the most appropriate transform.

Transforms are one way of modifying MLR's to model the non-linearity in the weather-demand relationship. Another class of methods are the switching regression models. A simple example is illustrated below.

$$y(t)_1 = a_0 + a_1 x_1(t) + a(t) \text{ where } x_1(t) < x_0(t) \quad \text{Equation 5.7a}$$

$$y(t)_2 = \beta_0 + \beta_1 \{x_t(t) - x_0(t)\} + b(t) \text{ where } x_1(t) \geq x_0(t) \quad \text{Equation 5.7b}$$

where

$x(t)$ = weather variable, which is the explanatory in this case

β_0 = equation 5.3

The switching regression model combines two MLR straight-line approximations each with a different gradient. A graphical representation is given in figure 5.7. The first model approximates the relationship for temperatures below the threshold, $x_0(t)$ whilst the second model estimates the relationship for temperatures greater than the threshold.

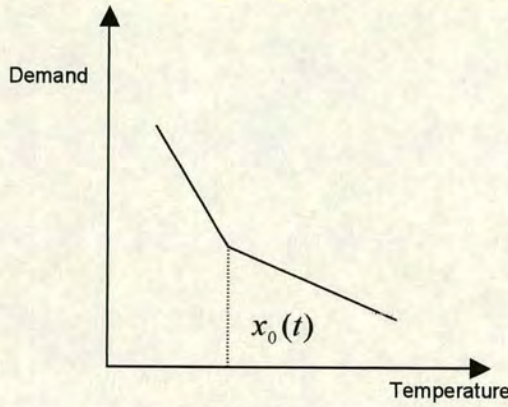


Figure 5.7. Switching regression model

This method gives a better approximation than the standard MLR method. It has two major drawbacks; Firstly, the switch between the models is sudden and might not be reflective of the relationship around the crossover point. Secondly, the crossover point is determined in practice through trial and error and will thus be influenced by the modeller's subjectivity.

5.5.1.2 Polynomial and trigonometric regressions

Though MLR is widely used, it has its limitations, which could be improved using transformations. Another class of transformations uses polynomials to describe the relationship between the explanatory variables and demand. The polynomial regression model is in the form:

$$y(t) = a_0 + a_1 x_1^j(t) + \dots + a_n x_n^k(t) + a(t) \quad \text{Equation 5.8}$$

where:

j and k = The regression polynomials

The major drawback with this method is that the data points close to the mean heavily influence the shape of the polynomial. This biases the forecast and provides unreliable descriptions of the behaviour at either ends of the explanatory variable range. For example at high or at low temperatures.

In some cases polynomial terms are combined with other nonlinear terms to obtain more accurate models. The most common approach involves combining with trigonometric terms to produce a parsimonious model⁷². These models are used to represent the periodicity within the seasonality, day-of-week and time-of-day cycles. An example with the two variable polynomial model is illustrated below:

$$y(t) = a_0 + \sum_{i=1}^d a_i x_i^i(t) + \sum_{j=1}^{\lambda} [\mu_j \cos\{jx(t)\} + \delta_j \sin\{jx(t)\}] + a(t) \quad \text{Equation 5.9}$$

where

$$\begin{aligned} d &= 2 \\ \lambda &= \text{a set of estimator parameters that minimize the risk function } R(\lambda) \end{aligned}$$

5.5.1.3 Non-parametric regression

This method has only recently been applied to demand forecasting because of its high computational requirements. Non-parametric regressions are a collection of techniques for performing regressions when there is little prior knowledge about its shape. Nonparametric methods allow a load forecast to be calculated directly from historical data without having to formulate a particular load model⁷⁴ as is the case with parametric models. The fundamental assumption of this method is that a load prediction is a conditional expectation of demand given the time, weather conditions, and other explanatory variables. This is calculated directly from historical data as a local average of observed past loads.

The local averages are found by using smoothers. A smoother is a tool for describing the trend in $y(t)$ as a function of the explanatory variables. Smoothers are useful because the amount of horizontal scatter in the data will often make it difficult to visualise the trend in a set of observations. The most commonly used smoothers are described below⁷⁵:

- **Running line:** This is the simplest smoother. Firstly, the data is divided into a moving 'window' of points. The window size or the 'size of neighbourhood' is typically 10-50% of the data. A simple linear regression is computed for each window. The windows overlap and are influenced by the results from adjacent neighbourhoods. The disadvantage of this method is that anomalous data points could influence the regression line.
- **Modified running line:** The modified running line smoother eliminates the effect of influential data points. This is done by using only a certain fraction of the data points in each neighbourhood, that seem to fit the regression line. For example, 70% of the data in a given neighbourhood might be used together with the subset of data points that minimise the residual sum of squares. Such computations are computer intensive. For both running line methods, it is considerably more difficult to obtain statistical inferences like t-statistics on the data. This is because of the absence of certain exact distributional results.
- **Kernel regression:** This is one of the most commonly used smoothing method. In kernel regression the size of the neighbourhood is called bandwidth, as with the previous methods, the neighbourhoods overlap. With the running line methods, points within each neighbourhood are equally weighted, with the kernel methods, the weighting is

dependent on the distributions around the local mean. This eliminates the influence of anomalous data points. Kernel estimators can perform badly at the boundaries of the predictor space and also when the true function is linear.

- **Local regression:** This is a combination of the running line smoother method and kernel methods. As in the running line method, a least squares fit is produced but as in kernel regression, weights are used that reflect each point's distance from the local mean. Therefore the fit in each neighbourhood is a weighted least squares fit. It has the advantages of kernel, in that; points far from the local mean have less influence on the regression line, eliminating bias. When the predictor has two or more dimensions, Kernel estimates are influenced by the boundary effect, this effect is eliminated with local regressions.
- **Splines:** They are generally defined as piecewise regressions, in which the curves (or line) are constructed individually and then pieced together. There are different types of splines, namely linear splines, splines with polynomial terms and smoothing splines. A smoothing spline differs considerably from the other previous smoother methods. In particular, rather than fitting lines and/or curve segments to the data, the smoothing spline uses optimisations. The major differences of splines are that: Firstly, the data segments fit disjointedly, whereas in local regression they overlap. Secondly, the computational complexity of extending the number of explanatory variables increases in the order of n^3 for each additional variable. This makes it unsuitable for electricity demand forecasting.

Nonparametric methods offer the regression considerably more flexibility than the parametric methods described. This is firstly because the shape is made up of a combination of separate lines or points, which would represent the relationship better. Secondly, the shape is more accurately representative of the data because it is inferred mathematically from the historical data not determined either through trial and error or by visual inspection, as is the case with parametric regressions.

Caution must be exercised when using these methods because of the potential danger of over fitting the data. If this happens the model will act as a lookup table and will be inappropriate for forecasting a wide array of situations. Added to this, the accuracy of a nonparametric method relies on the adequate representation of future conditions by historical data. The model will give poor predictions for events not present historically. Nonparametric regression will not be as efficient as parametric methods if a simpler model adequately defines the relationship.

5.5.2 Time series methods

Time series methods have been very popular within the ESI for the past 60 years. They have several advantages and disadvantages over regression models. Their principal advantage is their structural simplicity. The main disadvantage is that they generally do not describe a 'cause-and-effect' relationship, only one of dependency. Thus, these models cannot provide insight to the demand-weather relationship, fundamental to effective forecasting.

In time series models, electrical demand $y(t)$ is modelled as the output from a linear filter that has a random series input, known as white noise. This random input has a zero mean and an unknown fixed variance. A block diagram of the load time series model is shown in figure 5.8.



Figure 5.8. Load time series model

Depending on the characteristic of the linear filter, different models can be classified⁷⁶. The most common of these are described in the following sections.

5.5.2.1 Moving Average (MA) process

In this method the current value of the demand is expressed linearly in terms of the current and previous values of white noise. White noise is determined iteratively, as the forecast errors from previous load forecasts. The moving average (MA) is defined as

$$y(t) = a(t) - \theta_1 a(t-1) - \theta_2 a(t-2) - \dots - \theta_n a(t-n) \quad \text{Equation 5.10}$$

where

$$\begin{aligned} a(t) &= \text{white noise} \\ \theta &= \text{moving average coefficients} \end{aligned}$$

5.5.2.2 Autoregressive (AR) process

The load in this method is determined as a linear extrapolation from previous values. This gives an extrapolation from data sets that have been derived through historical correlation of dependent variables. The oldest previous value determines whether the series is positive or negative. It takes the form

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \dots + \phi_n y(t-n) + a(t) \quad \text{Equation 5.11}$$

where

$$\begin{aligned}
 y(t-1), \dots, y(t-n) &= \text{Previous demand values} \\
 a(t) &= \text{White noise} \\
 \phi_n &= \text{Autoregression coefficients}
 \end{aligned}$$

5.5.2.3 Autoregressive Moving-Average (ARMA) process

This combines both the autoregressive and the moving average models. It captures both the self correcting nature of the moving average method whilst expressing the demand as a function of previous demand. It takes the form

$$y(t) = \phi_1 y(t-1) + \dots + \phi_n y(t-n) + a(t) - \theta_1 a(t-1) - \dots - \theta_n a(t-n) \quad \text{Equation 5.12}$$

5.5.2.4 Autoregressive integrated moving average (ARIMA) process

The previous time series models are stationary processes. This means that the mean and covariance among its observations do not change with time. This is not the case with electricity demand because of the time-of-day, day-of-week and seasonality effects. The series has to be transformed to a non-stationary process, which is done using differential forms of AR and MA methods. Empirical research, notably Hagan and Behr⁷⁷ show that this approach is rather complicated, though its short-term results are very good.

5.5.2.5 Transfer Function (TF) modelling

The ARIMA model forecasts are essentially extrapolations of previous load history and are unable to cope adequately with weather effects. Transfer function models allow for the inclusion of some independent variables. The transfer function is obtained using one of a variety of estimation techniques. The extrapolation is done using one of the time series noise models described in the previous sections. Both outputs are summed up to give the forecast. The transfer function model is illustrated in figure 5.9.

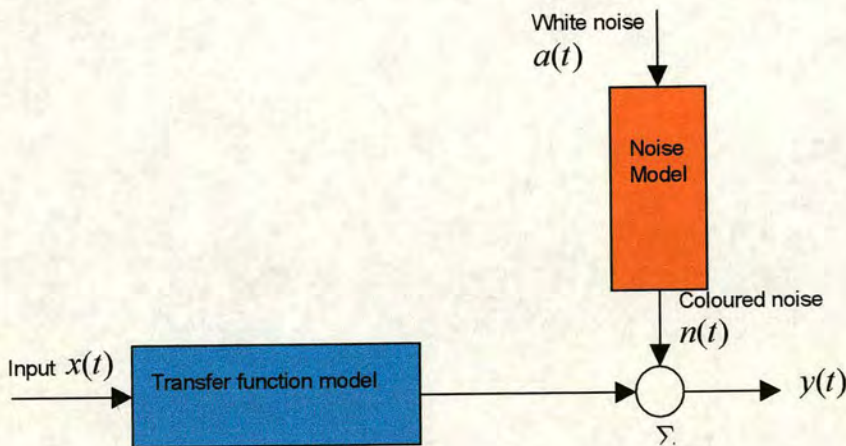


Figure 5.9. The transfer function model

This approach can provide very good forecasts as long as the independent variables are similar to those from which the transfer function is created. However, during forecasting because the independent variable is generated internally, any major deviations from the mean will cause inaccurate forecasts.

5.5.3 End-User models

In the previous methods forecasts are derived through extrapolations from historical data. In end-user models, they are calculated as the sum of electricity end-use in residential, commercial and industrial sectors. The basic structure of an end-use forecasting model is based on the following relationship.

*Energy consumption by end-use group = Number of users in group * Energy use per end-user*

Within the commercial/ domestic sector, end-user models begin by assuming that the total energy consumption equals the sum of the energy consumed for each end-use category. This assumption can be represented by the product of the following terms:

- The number of devices in each category
- A measure of the energy-using capacity of the device
- A measure of the average rate of utilisation of the device

This can be represented in the equation form

$$E = S \times N \times P \times H \quad \text{Equation 5.13}$$

where

E = Energy consumption of the appliance in kWh

S = Saturation in the number of such appliances per customer

N = Number of customers

P = Power required by appliance in kW

H = Hours of appliance use

Customer households are grouped into different types based on size, number of occupants, type of house, demographics, and socio economic group⁷⁸. Sample households are selected from each customer group and the load for each energy-consuming appliance is monitored over a period of time⁷⁹. The results are gathered to build the database for forecasting. The data base is modified

through regular monitoring and re-sampling the households. The major disadvantage with end-user models is that they are information intensive and ignore changes in consumer behaviour.

5.5.4 Artificial Intelligence methods

Artificial Intelligence (AI) has been around since the late 1950's. The advent of more powerful computers has seen their application soar over the past decade. Artificial Intelligence is defined as 'Computer processes that attempt to emulate the human thought processes that are associated with the activities that require the use of intelligence'⁸⁰. For the purposes of this discussion, Artificial Intelligence methods will be classed into four groups; Knowledge-based Systems, Fuzzy Logic, Artificial Neural Networks and Genetic Algorithms. These will be discussed in turn with specific focus on their application to short-term load forecasting.

5.5.4.1 Knowledge-based systems

These systems capture knowledge on a specific field in a computer interpretable model. There exist two subsets: Decision Support Systems (DSS) and Expert Systems (ES).

A decision support system allows the user to access previously collated knowledge bases. The knowledge base is in the form of a working computer model, or set of models. The user has access to some of the variables within the model and can alter them to create the environment in which the forecast is to be made. This is usually done through a questionnaire styled interface. The answers are used in the model along with internal, endogenous, data to provide either a solution or further questions.

The applicability of DSS systems to short-term forecasting has been very limited because models have to be very large to contain the information required. Large scale models suffer from feedback and non-linearity problems which makes models difficult to solve. DSS systems can be improved by incorporating other AI techniques.

An Expert System is similar to a Decision Support System except it uses a knowledge base of rules rather than a model to provide the solutions. The expert knowledge is stored as a number of strict technical instructions and rules of thumb that apply to the specific problem. In the case of short-term demand forecasting the analogical thinking that goes into the system operator's intuitive forecasting is reduced to logical steps and programmed⁸¹. Figure 5.10 shows the general structure of a typical ES

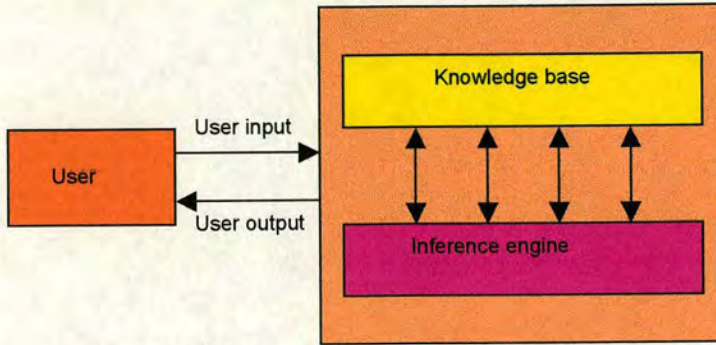


Figure 5.10. Structure of a typical expert system

The rules are normally divided into separate data bases, each concerning a particular discipline. They are then presented to the inference engine as - *If condition(s) then action(s)* - statements, written in plain text to facilitate modification and understanding. The inference engine is a separate programme that acts as an interface between the user and the knowledge base. This is the most complicated part of an Expert System.

Though Expert Systems have greater applicability in forecasting than Decision Support Systems, they are inefficient at coping with complex problems requiring a large rule base. Problems caused by large models are further accentuated in systems that contain a feedback loop, for example lagged variables in a forecasting model. This is due to the iterative methods used to approximate such loops. Another disadvantage with ES is that, expertise is difficult to extract from experts and can be highly heuristic. Heuristic rules can fail when given insufficient or unexpected data. ES often work very well when used in conjunction with appropriate inferencing methods⁸².

5.5.4.2 Fuzzy Logic

Traditional logic deals with precise values, yes or no, 1 or 0. These are known as 'crisp values'. Fuzzy Logic, on the other hand, attempts to model the impreciseness of human reasoning by representing uncertainty for the variables that are used by assignment of a 'set' of values to the variable. Each value has a 'degree of membership' of the set which represents the probability of the variable having that value⁸³. A 'membership function' is a curve that defines how each point, over the whole range of input values, is mapped to a membership value between 0 and 1. The membership value can also be viewed as the degree of belonging to a particular group or set.

Fuzzy logic can be applied to the short-term load forecasting problem because

- It is a multiple-input, multiple-output unknown dynamic system

- There is a high degree of periodicity within the load forecasts and fuzzy logic has been proved to have great capabilities in drawing similarities from large data sets⁸⁴.

Short-term forecasting is performed by pattern recognition. Fuzzy logic is used to classify and map the input and output pairs during training. To forecast, the fuzzy logic forecaster references its database, based on its input value then weights and sums the most likely output⁸⁵.

5.5.4.3 Artificial neural networks

The human brain is the most complex computing device known to man. Researchers have sought to create a computer model that matches the functionality of the brain's fundamental nerve units, neurons. The neuron is the base cellular unit of the nervous system, in particular, the brain. Each neuron is a simple micro processing unit which receives and combines signals from many other neurons through input structures called dendrites. If the combined signal is strong enough, it activates the firing of the neuron, which produces an output signal along a component of cell called the axon. The axon of a neuron splits up and connects to dendrites of other neurons through a junction known as a synapse. The transmission across the synapse is chemical in nature and the amount of signal transferred depends on the amount of chemicals or neuro transmitters. The quantity of neuro-transmitters present is what is modified when the brain learns. This combined with the processing of information in the neuron form the basic memory mechanism of the brain.

In artificial neural networks, the unit analogous to the biological neuron is referred to as a processing element (PE) or perceptrons⁸⁶. This has many input paths and combines, usually by a simple summation, the values of these input paths. This is given below.

$$I_j = \sum_i W_j X_i \quad \text{Equation 5.14}$$

The combined input is then modified by a transfer function. The output of this function, y_j is passed directly to the output of the perceptron.

$$y_j = f(I_j) \quad \text{Equation 5.15}$$

The transfer function has a threshold that only passes information if I_j reaches a certain level. This function can be a variety of shapes though sigmoid is the most common. A sigmoid function ensures that the output value, y_j increases smoothly from 0 to 1, not abruptly as in hard-limiting functions. Both functions are illustrated in figure 5.13. The sigmoid function allows the perceptron to represent ambiguity.

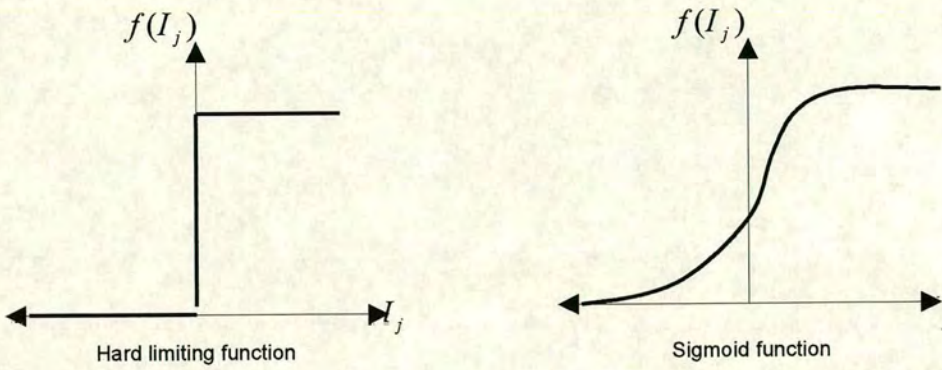


Figure 5.13. Perceptron transfer functions

The output path of the perceptron can be connected to the input paths of other perceptrons. These are weighted to correspond to the synaptic strength of the neural connections. Since each connection has a corresponding weight, the signals on the input lines will be modified by these weights prior to being summed up. Thus, the summation function is a weighted summation function. A model of a perceptron with weighted inputs is illustrated in figure 5.14.

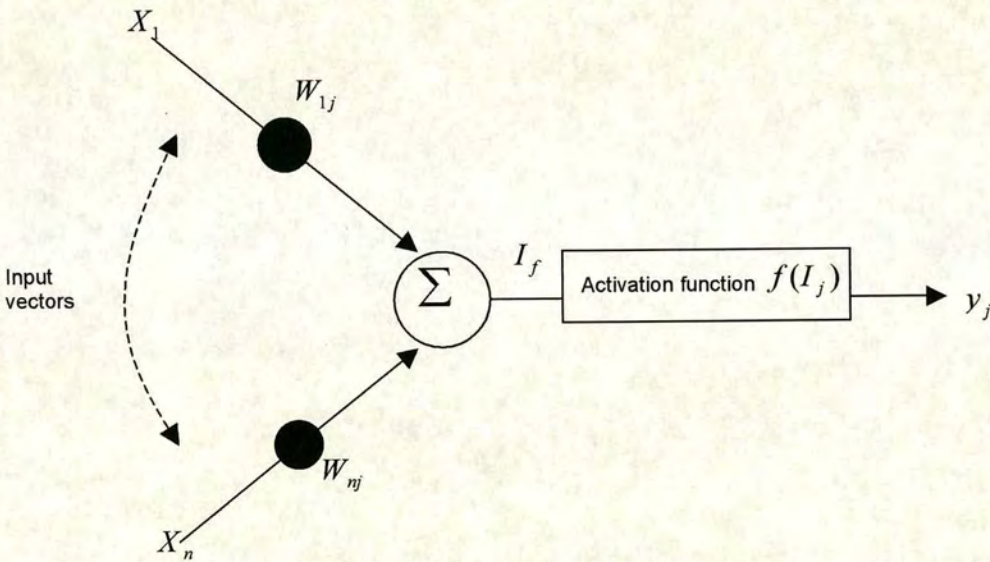


Figure 5.14. A model of a perceptron

A neural network consists of many perceptrons joined up in the manner described. The perceptrons are organised into groups known as layers. There are usually two layers connecting to the outside world: An input layer, where data is presented to the network, and an output layer, which holds the response of the network to a given input. Layers distinct from the input and output are called hidden layers.

Within a network, there are full or random connections between successive layers. The strength or weight of each connection is determined during the training process. During training, the network infers relationships by presenting an input and its desired response to the input and output of the network. This method of training is known as ‘supervised learning’. When no desired output is shown to the network, the learning is known as ‘unsupervised’. Regardless of the kind of learning used, an essential characteristic of any network is its learning rule. This specifies how weights adapt in response to a learning example. A diagram of a full network is shown in figure 5.15. The example below is a known as a three layered network because it consist of an input, a hidden and an output layer.

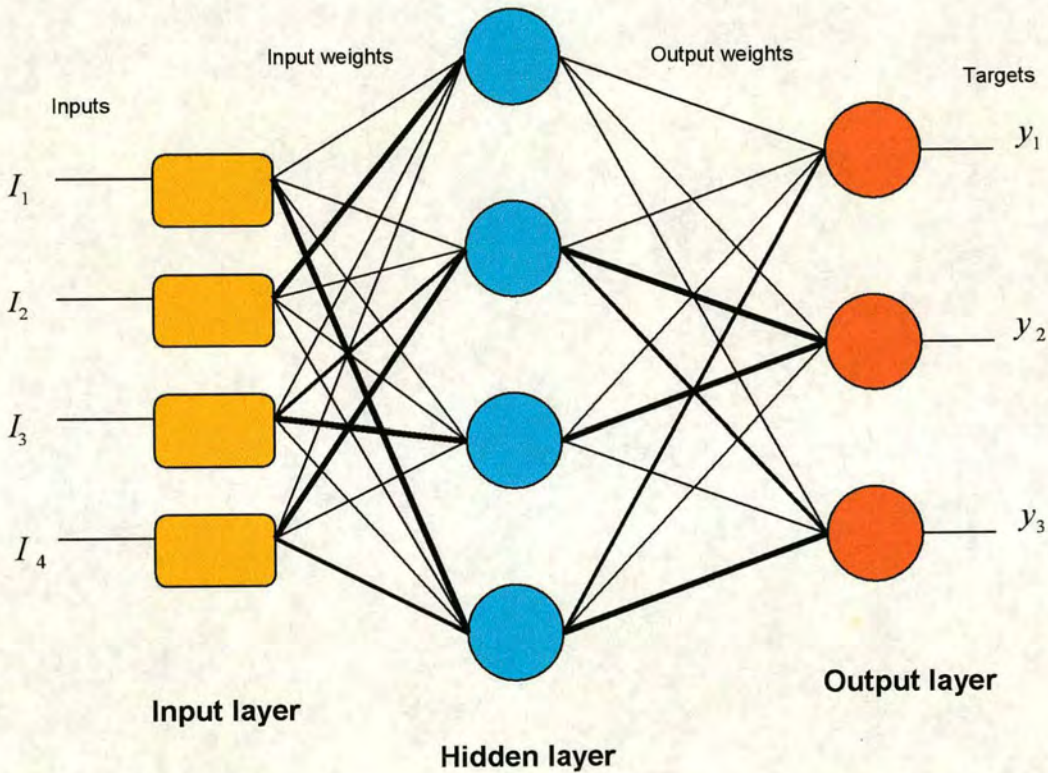


Figure 5.15. A three-layer artificial neural network

The neural network can be viewed as a black-boxed non-parametric regression. ANNs are the most widely used AI method in short-term forecasting⁸⁷ because they lend themselves intuitively this application. The explanatory variables are presented at the ANN’s inputs and the load at the output. The ANN infers relationships between the demand and the explanatory variables. Once trained the ANN is capable of forecasting from inputs outside the historical training space.

5.5.4.4 Hybrid systems

The final group of AI systems currently in use for short-term forecasting are hybrid systems. These combine two or more of the AI techniques described to perform the forecast. In the majority of these methods either a fuzzy logic or an expert system is used to classify the input based on several characteristics such as season, day and time⁸⁸. More recently, fuzzy classifiers have emerged as the most prevalent classification method because of their ability to handle ambiguous data. Added to this, their implementation does not require a great amount of expert knowledge. After classification, inputs are presented to an ANN via an interface for relationship inference.

Another hybrid method is the combined Genetic Algorithm (GA) and Artificial Neural Network approach⁸⁹. GA's are essentially an optimisation method which models the biological process of natural selection to evolve generations of offspring that are closer to the optimum than their parents. They are used to select the optimal training inputs and outputs for the ANN.

Though their conception and construction is more complex, hybrid systems are becoming increasingly popular in academic literature. Some of the ANN-Fuzzy hybrid systems have reported superior performance over other methods. The neural network is able to draw more precise relationships because classified data is very specific. The major drawback of hybrid systems is that they require a much larger database to supply the ANN with enough training data for each class.

5.6 Comparison of the forecasting methods

The various short-term forecasting methods have been presented and discussed in the preceding sections. This section argues the merits and demerits of each generic group with a view to narrowing the selection of methods for analytical comparison in chapter 7. The appropriate choices for Supplier's short-term forecasting requirements will be made from the following discussion.

5.6.1 Time series models

These models employ historical data for extrapolation to obtain future hourly loads. The disadvantage of these models is that they model a stationary load trend. Even after incorporating transfer functions, the weather and any other dynamic factors that contribute to load behaviour are not fully utilised within the forecast. This could lead to poor forecasts during dynamic periods, for example, when the weather is heavily influencing the load.

5.6.2 Regression methods

Regression methods analyse the relationship between the load and other influential variables. Linear regressions have the disadvantage of representing the relationship linearly. With transformations, the nonlinearities are better represented, however, selecting the most appropriate transformation is done visually, which could be deceptive. Added to this, the curve is heavily influenced by data close to the mean. Non-parametric regression provides the most appropriate fit however, this requires complex modelling techniques and heavy computational effort.

5.6.3 Artificial intelligence

Knowledge based systems model the knowledge of a human expert to develop the rules for forecasting. Transforming the knowledge of an expert to a set of mathematical rules is often very difficult. Fuzzy logic systems produce much better forecasts but there are difficulties when forecasting beyond the boundaries of the training data set. This is because forecasts are based on a discrete set of historical conditions.

ANN methods are able to learn and extract complex relationships from multivariate data. Their main disadvantage is that they provide a black box solution, as such it is very difficult to assess the effect of individual variables.

Hybrid systems usually combine the advantages of classification with the ANN's ability to infer complex relationships. Their major drawback is that they require sophisticated model interfaces and are far more data intensive than the previous methods.

5.6.4 Model choice

The preceding sections discussed the various short-term forecasting methods with a view to selecting the most appropriate models for analytic assessment in the next chapter. To meet Suppliers forecasting requirements under NETA, the primary requirements governing the model choice are:

- 4-hour lead time
- Very high accuracy
- Flexibility
- Easy implementation.

Short-term demand is heavily influenced by independent dynamic factors like the weather. These are essential inputs into Suppliers forecasting models. Time-series models will not be adequate

because they do not take these factors into account. The inclusion of a transfer function will improve the forecast though not adequately for Suppliers requirements.

End-user models were considered because they use a bottom-up approach. This is different from the top-down approach used by all the other models. Another unique feature of this family is that they do not forecast from historical data. Unfortunately due to insufficient data, accurate implementation of an end-user model was not possible.

Over the past decade, the majority of the forecasting research and literature has focussed on Artificial Intelligence methods in particular, Artificial Neural Networks. ANNs are chosen for the analysis because of the following:

- They are most widely used ANN method
- They enable a multivariate implementation of the relationship
- The process infers a non-parametric regression between the explanatory variables and the load

Fuzzy logic forecasters were also considered however, they require large-scale models and are discrete resulting in reduced accuracy outside the boundary of the training space. ANN-Fuzzy hybrid systems would be very effective, but due to limited data they are omitted.

Regression models have the advantage of transparency and are capable of representing the dynamic factors within the forecasting model. The different types of regression were considered. It was decided to opt for MLR for the comparative analysis to determine whether a simple straight-line approximation would be good enough to predict demand at the BSP. Additionally they are the most widely used short-term forecasting method. Transformations were considered, however it was decided that as ANNs would be used in the analysis, mathematical transformation of the MLR would not represent a distinctly different class to warrant separate comparison. Non-parametric regressions are a unique class with different derivation from other regression methods. Its flexible regression shape is potentially more mathematically representative of the relationship than parametric models. However, they are not selected because neural networks perform non-parametric regressions and are more easily implemented.

5.7 Summary

The price of volumetric risk under NETA could seriously damage Suppliers financially and, as a consequence they will need to have demand forecasting methods in place to provide the system operator with FPN's each half-hour for every BSP they operate. The accuracy of the forecasts will aid risk management activities and determine the exposure to imbalance prices.

This chapter focussed on the subject of accurate demand forecasting from a Suppliers perspective with a view to selecting the most appropriate forecasting methods for comparative analysis. Firstly, the fundamental factors affecting demand: economic activity; time factors; weather and random influences were discussed. This was because they have a large effect on demand and are as such essential inputs into any forecast model.

This was followed by a description of forecasting time horizons: short; medium and long; and the fundamental demand drivers over each horizon. To facilitate the choice of the most appropriate short-term forecasting method, a detailed description, followed by a discussion of the most commonly used forecasting methods was presented. The generic classes considered were; regressions, time-series, end-user and artificial intelligence methods.

From the discussion, Multiple Linear Regressions and Artificial Neural Networks were chosen as the most appropriate for comparative analysis in the next chapter.

6. Data issues and selection of forecast inputs

6.1 Overview

It is essential that Suppliers and other market participants choose the most effective short-term forecasting method because of the value of volumetric risk brought on by New Electricity Trading Arrangements (NETA). In the previous Chapter, electricity demand forecasting was discussed and the complexities involved were presented. Two short-term forecasting models, Artificial Neural Networks (ANN) and Multiple Linear Regressions (MLR) were chosen for the comparison. A forecasting model can only aspire to be as good as the quality of the data from which it infers relationships and predicts demand.

In this chapter the data issues involved in forecasting demand at BSP level are discussed. SHE, the first-tier Supplier in the area considered in this study, is the largest user of demand-side management in the UK and this heavily biases the data. This bias has serious implications for any forecasting model. This chapter discusses the data issues involved in modelling and forecasting demand at a BSP in SHE's region.

The chapter starts with a discussion of the data requirements of the model. This is followed by an introduction and discussion of the teleswitch system used by SHE and the implications to short-term forecasting. The forecast configuration of the models is discussed next, followed by a statistical analysis of the data to select the most relevant inputs for the forecasting model.

6.2 Data Selection

As was mentioned in previous sections, domestic demand causes the greatest variability within Suppliers demand portfolios, so this demand segment represents the highest source of volumetric risk. To examine fully the limitations of the present techniques, it was decided to select a BSP supplying a predominantly domestic load, since the higher demand variation will pose greater challenges to the model's ability to infer relationships.

6.2.1 Demand data

Several locations within SHE's first tier supply area were considered. The Grampian region was chosen for closer examination, because it contained majority of SHE's domestic load in the Aberdeenshire area. Efficient half-hourly metering at the BSP level had only started in most

places after 1995 but even with this, there were still large gaps in the data sets of most of the Bulk Supply Points (BSP) observed because the majority of the metering systems needed time to bed-in.

The 33kV BSP at Dyce, near the Aberdeen airport was chosen because it had a robust data set and supplied a predominantly domestic load of some 13131 customers. The magnitude of the load at this BSP had stayed the same over the preceding four years, implying there had been no significant load growths or reductions. As the data set was fairly stable, inferences about the future demand based on historic trends and cycles would be valid. Half-hourly metered data at this BSP was collected from 1st January 1995 till 31st March 1999.

6.2.2 Weather data

Due to the strong reliance of demand on the weather, it was decided to use hourly temperature and wind-speed and direction since they have the most tangible effect on demand. These were purchased from the Meteorological Office in Glasgow for the period from 1st January 1995 till 31st March 1999. The data was measured at the Dyce weather station near Aberdeen airport.

The Dyce weather station was selected because it was close to the BSP and because the variables at this station gave an accurate measurement of the weather variables affecting the demand at the BSP. Research shows that the further away the weather station from the demand, the lower the influence of the measured variables on demand. Notably, the influence of wind speed measurements on demand is greatly reduced with increasing proximity of the measuring station to the demand. This is often due to many factors including the effect of numerous barriers such as buildings and trees.

6.2.3 Time horizon

The time horizon over which the models are trained and used to perform the forecast depended on firstly, data availability and secondly, on the relevance of the information contained within the data. A 3.5-year horizon was chosen because it was long enough to contain any possible trends and short enough to exclude obsolete ones.

6.2.4 Software choice

Computational power was needed to build the forecast models and to perform the statistical and mathematical analysis required for the input data selection and result comparisons. The most appropriate software for the project was chosen to meet the following computational requirements:

- Ability to handle and manipulate large time series database
- Strong programming capabilities for modelling
- Interfacing capabilities with C
- Speed
- Good library of statistical and mathematical functions
- Powerful graphics

Version 5.3 of MATLAB by Mathworks, was selected.

MATLAB is a language based on C and FORTRAN. Its strongest advantage is its ability to perform advanced matrix manipulations using very simple commands. There are hundreds of predefined commands and functions that can be enlarged by user-defined functions. Such features were very desirable, because the demand and weather time series data were compiled as large matrices. MATLAB also has very powerful graphing and visualisation capabilities that is useful for visual comparison and result presentations and it can be interfaced with other programmes. FORTRAN and C routines can be called from MATLAB and vice versa, increasing its power and flexibility.

6.2.4.1 Artificial Neural Network

There are numerous ANN packages available with their own advantages and disadvantages. Several such packages were compared.

- Neural Works: This package is produced by NeuralWare, it is very user friendly and has good graphics. Its major disadvantage is that it is slow and rather cumbersome
- Neural Network toolbox: This is part of the MATLAB package. It is a well-designed toolkit with good instructions, it is faster and more powerful than Neural Works, and has more features. Its major drawbacks are slow speed and difficulty of implementation.
- In house MLP: Dr Peter Edwards, a former PhD student and Lecturer at Edinburgh university, designed this ANN. It was written in C for the UNIX network. Though it is not as user friendly as the previous networks, it is faster and more powerful than the other packages and has very few commands.

It was decided to use Dr Edwards Neural Network because it was much faster and more stable than the other methods considered. Added to this, Dr Peter Edwards was at hand to give advice and to offer suggestions.

6.3 Characteristics of the load

Until recently SHE was a vertically integrated entity; generation, transmission, distribution and supply were all managed centrally. The most efficient way to operate a vertically integrated entity is to tailor supply to optimise generation. This allows generation to operate with a high load factor but requires demand to be shifted into the daily troughs.

The majority of SHE customers use electricity as their main source of space and water heating. This demand makes up the majority of the domestic consumption with, in some cases, over 75% of the total demand. For the reasons given above, SHE had decided to adopt 'valley filling', i.e., to shift heating demand to fill the demand troughs. These troughs occurred overnight, midmorning and early afternoon.

The demand side management was achieved by radio teleswitching.

6.3.1 Radio Teleswitching

Radio teleswitching, where load control is achieved via a radio-controlled switch on the customer's premises is the most commonly used method of demand side management in the UK and SHE is the biggest user of this system.

After extensive trials the Electricity Association approved the radio teleswitching system in 1979⁹⁰. In operation, Regional Electricity Companies (REC) submit their load control request to the Central Teleswitch Control Unit (CTCU), situated at the UK National Control Centre which in turn passes these messages on to the BBC Message Assembler (MA) at Broadcasting house. The data is then transmitted on the BBC radio 4 service at 198 kHz via transmitters at Droitwich Westerglen and Burghead and is received by the radio teleswitch on the premises of the customer.

The radio teleswitch is a simple low cost device with four contactors to control the customer's total electricity consumption. It consists of a 198kHz radio receiver, decoder, memory and four switches, each allowing up to 4 switch ON-OFF times over a 24-hour period. The contactors are connected to the domestic electrical circuitry as follows:

- Contactors A and B are rated between 2 and 4A and determine the charging rate for all electricity consumed within the premises apart from water and space heating.
- Contactor C is rated 25A and controls the water-heating appliance.
- Contactor D is rated 80/100A and controls the space heating.

Two types of user messages are sent down the teleswitch

- Programme instructions: These designate the switching times and the duration and are set up to 24 hours in advance. Once programmed, the system will operate on that regime until instructed otherwise. The programme instruction each teleswitch receives is determined in accordance with the customer's tariff structure.
- Command instructions take immediate effect and will override the programme message. This instruction is not remembered by the receiver and is cancelled when
 1. The programme reaches a switching boundary for contactor C and D or
 2. Another immediate command is received or
 3. 24 hours have elapsed since the previous instruction.

6.3.2 Dynamic load control

For load control, SHE's customers are divided into four geographical areas determined by postcode. Each geographical zone has its own switch ON-OFF times for customers on dynamic heating tariffs. The purpose of having four zones is to stagger the heating demand across the total system's demand troughs. The geographical zones are as follows

- Central highlands
- West and Southern highlands, Islands and Dundee
- North East
- Caithness, Sutherland, Orkney and Shetland

Each demand group has a separate water and space-heating regime with different switch ON and OFF times. Having received day-ahead temperature, wind speed and snow conditions from the Meteorological office, the loading engineer refers to a look-up table to determine the duration of the day-ahead heating blocks. The switch ON and OFF times for the day-ahead blocks are sent out 24 hours ahead via Broadcasting House to the individual teleswitches. In the case of unexpected circumstances, immediate messages can be sent a few minutes in advance to over-ride the original message.

6.3.3 Customer heating tariffs

For the purpose of this discussion, only customers in the North-eastern demand zone will be considered. SHE offers each of their domestic customers a choice of heating tariffs based on price. These tariffs are outlined below.

- **Total Heat Total Control (THTC):** On this tariff, customer water and storage heaters are dynamically controlled. There are 3 daily space heating blocks, and two water-heating blocks. The switch ON-OFF times for the water heating are 04:15 to 07:45 and 13:30 to 15:30, whereas the duration of the space heating blocks are determined by the forecast day ahead temperature. The switch ON-OFF times for customers on this tariff in North-Eastern Scotland are shown in Table 6.1.
- All other heating appliances, for example convector heaters can be connected to, and served by the space heating contactor.
- **Economy tariff:** Customers are supplied a fixed 8-hour space heating block regardless of the temperature. The heating block is from 23:00 to 07:00 in the winter and from 00:00 and 08:00 in the summer.
- **Restricted Heating Tariff (RHT):** This is a hybrid of the two systems. The customer is offered the economy tariff and a choice of one of the THTC heating blocks.

Temperature range	First block		Second block		Third block		Total supply hours
	ON	OFF	ON	OFF	ON	OFF	
< -7	00:00	06:00	10:30	13:30	18:00	21:00	12
-6	00:00	06:00	10:30	13:30	19:00	21:00	11
-5 to -4	01:00	06:00	10:30	13:30	19:00	21:00	10
-3 to -2	01:00	05:00	10:30	13:30	19:00	21:00	9
-1 to 1	01:00	04:30	10:30	13:00	19:00	21:00	8
2 to 4	01:00	04:30	10:30	12:30	19:00	21:00	7.5
5 to 7	01:00	04:00	10:30	12:30	18:30	20:30	7
8 to 10	01:00	04:00	10:30	12:00	18:30	20:00	6
7 to 11	01:00	04:00	10:30	11:30	18:30	19:30	5

Table 6.1. Total Heat Total Control space heating switch ON-OFF times for customers in North Eastern Scotland.

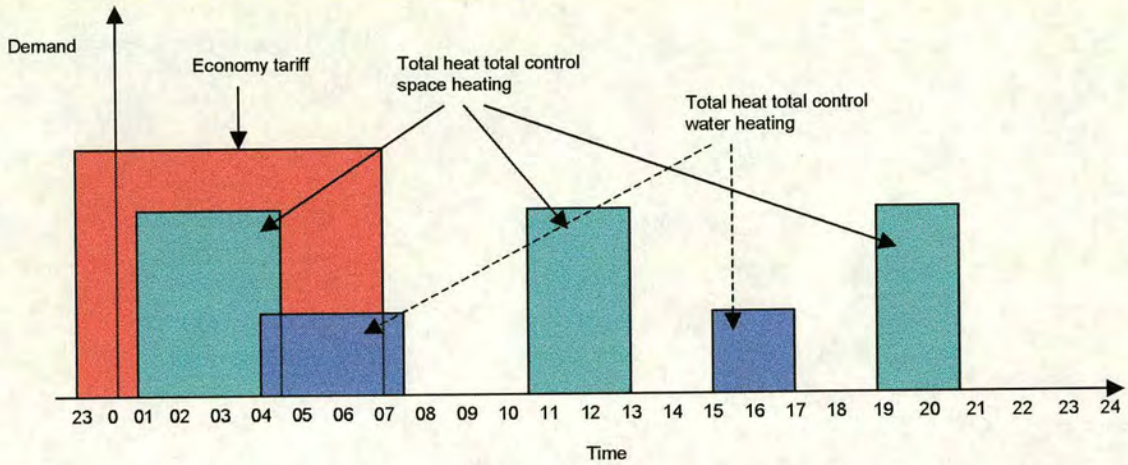


Figure 6.1.Example of daily switch ON-OFF times

It is impossible to determine with any certainty the numbers of customers on each tariff at a given time. This is because customers change tariffs frequently and information about the customer numbers on each tariff is not accessible. Figure 6.1 gives an example of the switch ON-OFF times of the heating demand in the Grampian region. The example uses a forecast minimum day-ahead temperature of 1°C. The heights of the heating blocks in the graph are approximations.

6.3.4.1 Real-time example of the effect of switch ON, on BSP demand

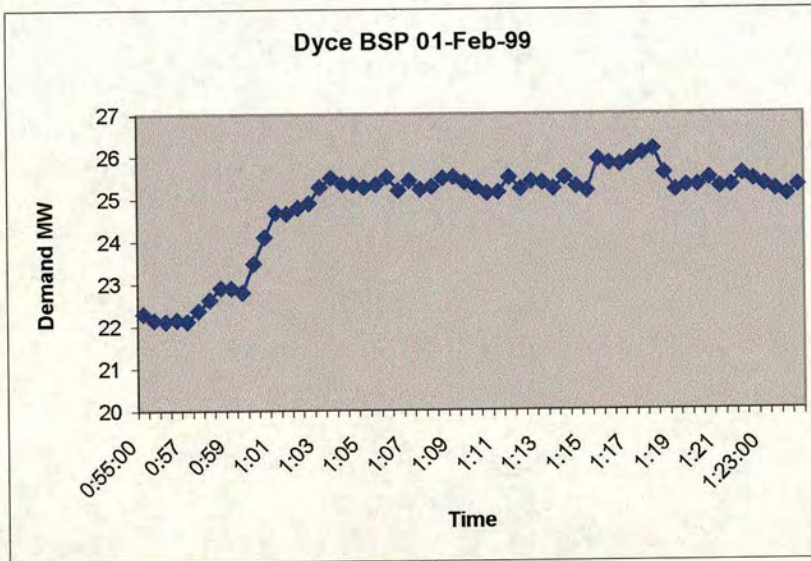


Figure 6.2. Real-time demand at BSP showing the ramping up of the first THTC demand block

The teleswitch operating system allows a switch closing time to be specified 8 times each hour as follows; on the hour or 7.5, 15, 22.5, 30, 37.5, 45 or 52.5 minutes past the hour. Each individual

teleswitch applies a random offset of between -3.5 and +3.5 minutes to the programmed opening and closing time. This causes a 7 minute rising and decaying ramp when the teleswitch Group is switched ON or OFF, effectively dampening the effect on the system. A real-time example is shown in figure 6.2. The graph shows the real-time demand at the BSP during switch ON for the first THTC demand block on 1st February 1999. The minimum forecast day-ahead temperature was 1°C corresponding to a switch ON time at 01:00. Due to the random -3.5 and +3.5 minute offset, the ramp starts at 00:57 and ends at 01:03:30. In this example, customers on this demand tariff added an extra 4MW to the total demand.

6.3.5 Complexities of forecasting teleswitching demand

The extensive use of teleswitching adds additional nonlinearity into the demand data set, biasing the data and presenting added complexity to the forecasting model. The extra nonlinearity arises from the following sources.

- Heating demand makes up to 75% of the domestic customer demand. As the heating demand has been shifted to specific half-hours, the temperature and wind-speed demand relationship will be particularly strong across these half-hours. This increases the variations of the weather-demand relationship across the daily 24 hours, by confining the majority of the heating demand to specific half-hours. For example, the weather demand sensitivity is greater overnight because most of the heating demand is shifted to this period. This phenomenon is discussed further in the section on model input selection.
- Another source of nonlinearity occurs across the marginal half-hours. These are the half hours that are not always covered by THTC or RHT. For example, the half-hour between 12:30 and 13:00. This is covered when the forecast day-ahead temperature is less than 1° C. If there is a one-degree increase to 2° C, this half-hour is not covered. This anomaly will be very difficult to model within the forecast. There are 8 such marginal half-hours.
- When the forecast wind speed is greater than force 4, equivalent to between 5.66 metres per second (m/s) and 8.23 m/s, two degrees are deducted from the forecast day-ahead temperature. For example, if the forecast day-ahead temperature is 3°C, equivalent to 7.5 hours THTC and RHT daily teleswitched demand with moderate winds of 7.72 m/s forecast, this will be interpreted as 1°C translating into 8 hours heating demand. Ordinarily, the wind speed would be expected to increase the weather demand relationship, thus increasing the height of the demand block. However in the above example, the height as well as the duration of the heating block would increase.

- The height of the teleswitched blocks is not linearly proportional to the temperature and wind-speed, it depends on the factors outlined in section 5.3.4. Added to this, customers constantly change tariff, this adds a random element to the numbers on the different tariffs hence the numbers of customers switched on the heating blocks.
- The forecast day ahead temperature determines the duration of the heating blocks. Any errors in the forecast will mean that the heating period will not correspond to the actual temperature. For example, the weather models on which the Meteorological Office bases its forecasts do not have a feedback mechanism to recalibrate the model. Over time, the weather forecasts could become biased, increasing the likelihood that the teleswitched heating duration will not correspond to the actual temperature. Forecast models use historical demand and weather data to infer relationships and not the historical forecast temperatures on which the teleswitch regime is based. This indicates that even if the forecast models are able to model teleswitching effectively, it may be different to the actual switching regime on the day.

From the discussion above, teleswitching adds a high degree of nonlinearity to the relationship between the demand and its explanatory variables. This will present the forecasting model with the arduous task of inferring robust relationships from contaminated data to perform accurate predictions. So far, attempts to predict accurately teleswitched demand have been unsuccessful.

Demand forecasting at SHE's BSPs will pose greater challenges to forecasting than the majority of the BSP's in the UK and these should expose any limitations in the modelling ability of the chosen models, MLR and ANN.

6.4 Forecast configuration

The problem of short-term demand forecasting is far too complex to be solved by a single model. The two most commonly used methods either assign a model to learn and predict the demand pattern for each day of the week or each half-hour or hour of the day. For the purposes of this discussion these are known as Serial or linear and parallel forecasting respectively.

6.4.1 Parallel forecasts

In this method forecasts are performed according to the time of day. Every half-hour of the day is assigned an individual forecast model to learn and predict the demand in the specified half-hour. This method uses 48 separate models to learn and predict the demand.

The implicit assumption in this method is that the time of the day variations are much greater than the day of week variations. By presenting each model with a specific half-hour, the model is

able to extrapolate the intricacies of the half-hour without noise from other half-hours. Parallel forecasting requires 48 models for each half-hour of the day. An example of the model configuration is given in figure 6.3.

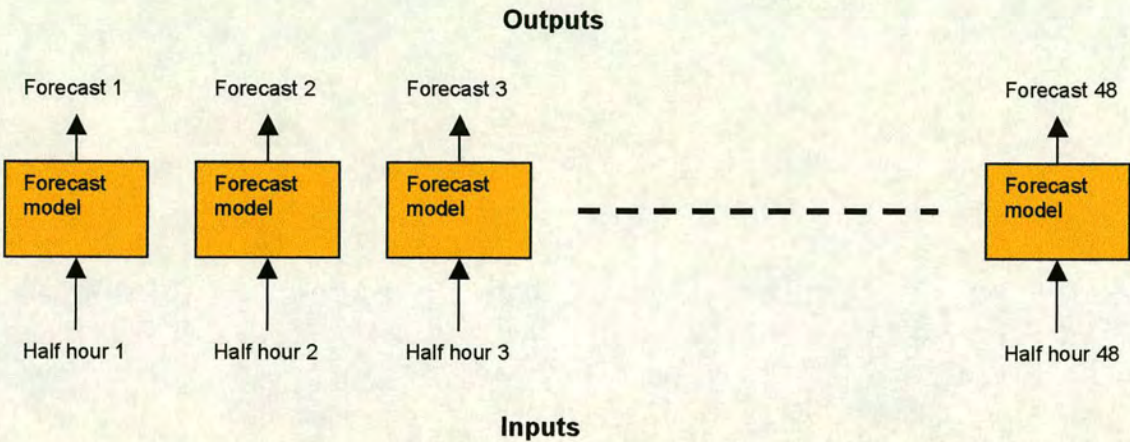


Figure 6.3 Schematic diagram of a parallel forecasting model

6.4.2 Serial (Linear) forecasting

With serial or linear forecasting it is assumed that the day of week variations are greater than the time of day variations. Forecasting is done according to the day of the week. A model is assigned for each day of the week. It is trained and used to forecast demand on the specified day. Serial forecasting requires seven models for each day of the week. An example of a serial forecasting model is illustrated in figure 6.4.

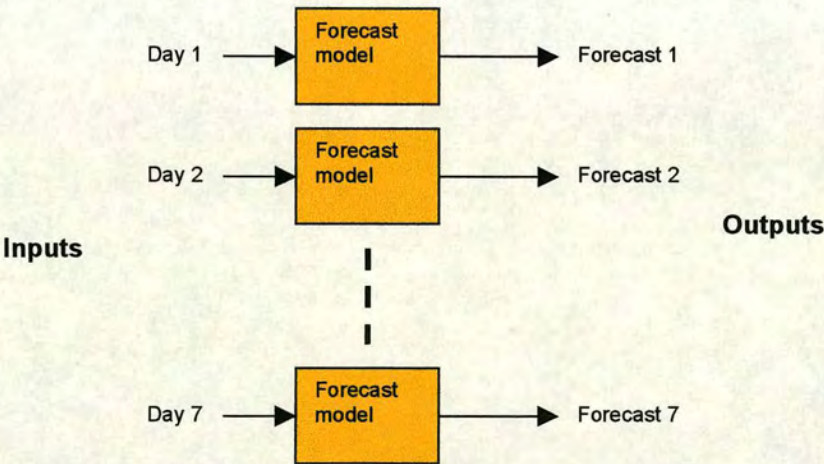


Figure 6.4 Schematic diagram of a serial forecasting model

6.4.3 Choice of the most appropriate configuration

For the purpose of the analytical comparison between the Multiple Linear Regression model and Artificial Neural Network models, it was decided to use a parallel forecasting configuration. This is because,

- Demand Side management, notably teleswitching, used extensively by SHE, introduces a high degree of nonlinearity. The nonlinearity is confined to specific half-hours, because teleswitching is used across specified half-hours of the day. This will increase the time of the day variation because the teleswitched half-hour will contain most of the weather sensitive load. The characteristics and demand drivers over these half-hours will be very different from those without heating demand. Added to this, the marginal teleswitched half-hours will present additional modelling challenges that will require the focussed application of an individual model to extrapolate and perform accurate predictions. Forecast models generalise from the data they are presented and it is assumed that they will make better inference from a purer data set, which parallel methods provide.
- The MLR represents its extrapolations as a straight line, as there are substantial time of day variations within the data set, a single MLR might have difficulty making meaningful inferences from which to base its forecast. With parallel forecasting the model should be capable of making much better extrapolations because the day is divided into 48 segments. There will be 48 half-hourly regression lines for each day. The combination of the individual half hourly regression lines could be viewed conceptually as a continuous flexible nonlinear regression line.

The assumption that the parallel forecast is the more appropriate configuration is investigated in later chapters, but for the purpose of the comparison, this assumption was adopted.

As parallel forecasting is the preferred model configuration, day-of-week identifiers were included in the models so as to classify the day of week variation within the data set.

6.5 Choice of appropriate inputs for the forecast model

In forecasting, the models considered learn to predict the future by drawing inferences from a training data set, usually historical data. The accuracy of the predictions is based upon the quality of the data presented to the forecast model so the choice of inputs from which the model learns, plays a crucial role in determining the accuracy of the model.

In this section statistical analysis is performed on the variables known to influence demand, also referred to as explanatory variables, to determine the most efficient predictors of demand. As stated in earlier chapters, the fundamental demand drivers are the weather, time variables and random factors. These will be considered as explanatory variables. Additional information may be contained in previous demand data and this also is considered.

Correlation analyses as well as graphical observations are used to perform the statistical analysis in this section.

6.5.1 Correlation analysis

Correlation is used to describe the strength of a linear relationship between two variables, by computing the correlation coefficient. The correlation coefficient has the following properties

- Its value is computed as a number between +1 and -1, a positive correlation coefficient implies a direct relationship while a negative coefficient implies an inverse relationship
- Values of +1 and -1 signify an exact direct and inverse relationship between the variables. A correlation of zero indicates the absence of a linear relationship between the two variables.
- The correlation coefficient measures only the strength of a linear relationship so a correlation of zero implies no linear relationship but not the absence of any other relationship.

Correlation analysis was performed between the demand and its most likely predictors, previous demand and weather variables. The results from this analysis were used to determine the most appropriate inputs to the forecast model.

6.5.2 Statistical analysis on weather variables

From the previous discussions, it was suggested that there is a strong inverse relationship between the demand and the temperature. This is confirmed in Figure 6.5, which shows a scatter plot of the average daily demand against average daily temperature. There are two distinct and separate clusters: encircled in green, the lower less dense cluster represents the weekend demand: encircled in red, the higher cluster shows the weekdays.

From the graph, the weekday demand is of greater magnitude, on average, than the weekend demand. This is probably because of the inclusion of the industrial and commercial load.

A simple linear regression was computed separately for the weekday and weekend demand using the temperature as the only explanatory variable, to determine whether the temperature-demand relationship is affected by the day of the week. The regression equation is presented below.

$$y(t) = a_0 + a_1 x(t) \quad \text{Equation 6.1}$$

- $y(t)$ = Electrical demand at time t
- $x_1(t)$ = Temperature variable correlated with demand
- $a_0(t)$ = Intercept term
- a_1 = Regression coefficient

During working days, the coefficients of the regression line were found to be

$$y(t) = 32.1138 - 0.6194x(t)$$

and at the weekends it was found to be

$$y(t) = 24.9672 - 0.559x(t)$$

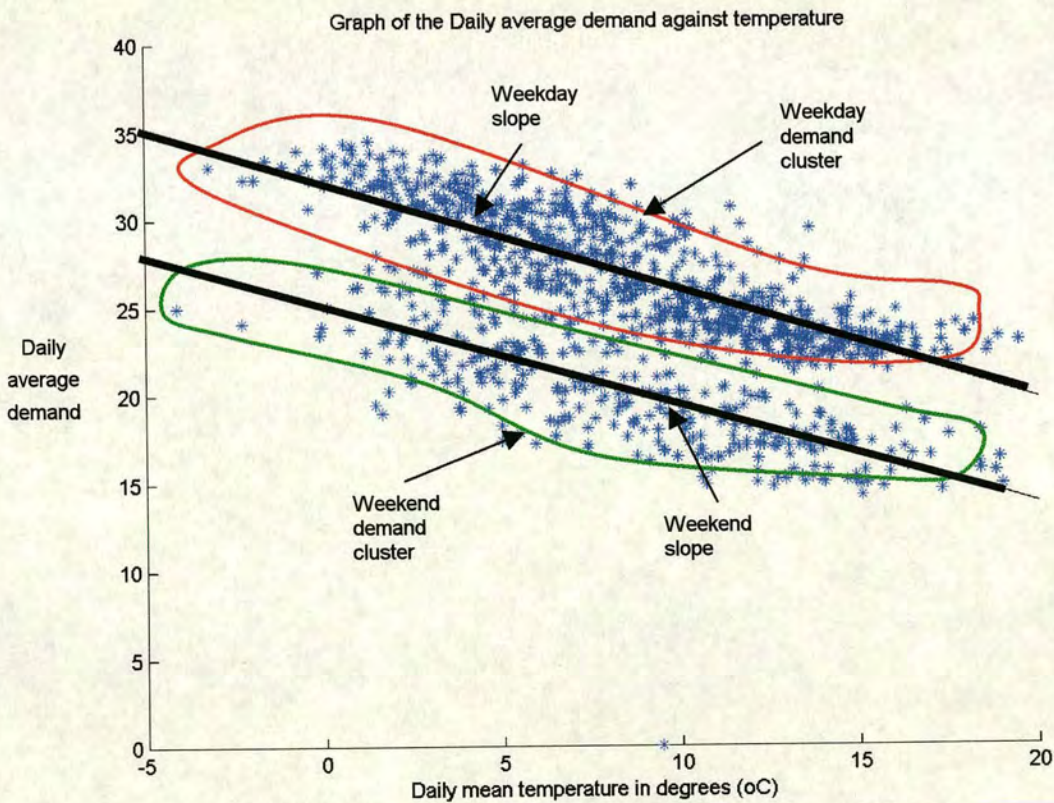


Figure 6.5. Scatter plot of the average daily temperature against the average daily demand

The slope for the weekday demand is slightly steeper than that of the weekend demand, implying that the temperature-demand relationship at weekends is different to that in weekdays. The relationship is less steep at the weekends, indicating marginally less sensitivity.

Identifying the reasons for the difference in relationship between weekdays and weekends is beyond the scope of the present work and the analysis is by no means a conclusive representation of the temperature-demand relationship, only of the linear component. However, it is apparent that the lines suggest a day of week effect and this will be investigated more thoroughly in the next section.

From the discussions in section 5.3.4 wind speed is expected to have an effect on demand. Figure 6.6 shows the average daily wind speed against the average daily demand: wind direction was not used in the analysis because of the added complexity.

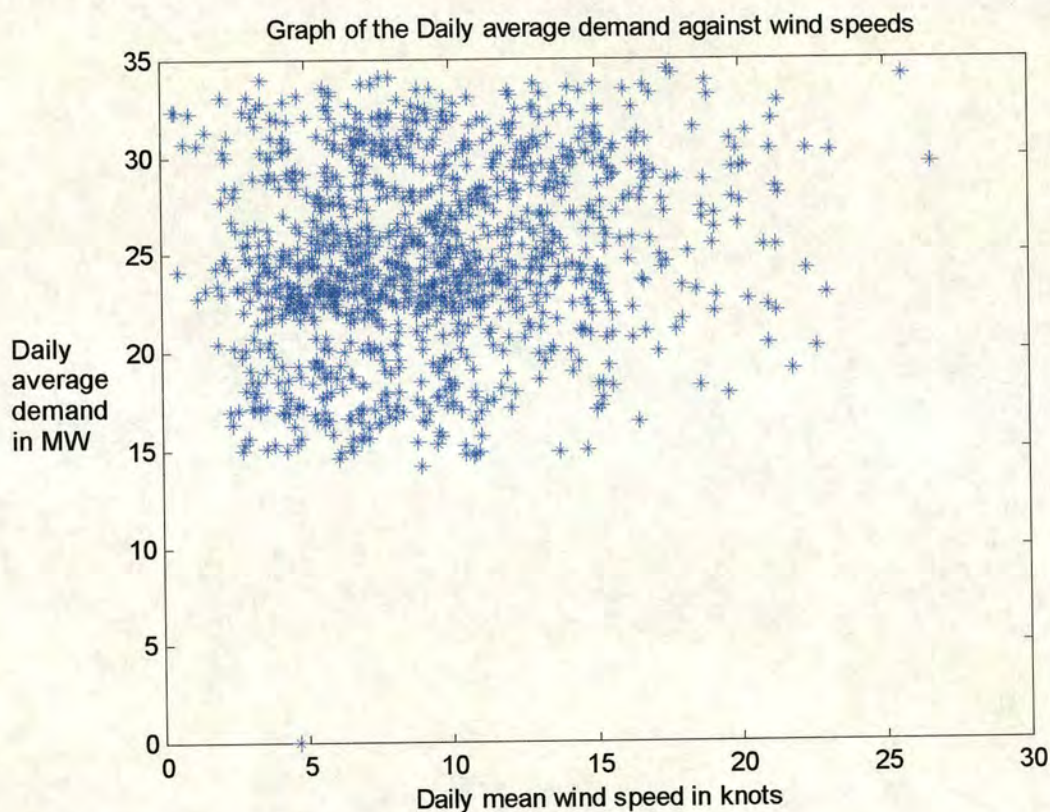


Figure 6.6. Scatter plot of the average daily wind speed against the average daily demand

Unlike the temperature-demand scatter plot, there is no noticeable pattern or relationship between the average daily wind speed and demand but it should be remembered that absence of a recognisable pattern does not imply the absence of a relationship with demand. Indeed, experience is that high wind speeds have a marked effect on demand, when combined with low temperatures. This is possibly because high wind speed increases rate of the airflow around

dwelling, which increases the rate of heat loss so more energy is consumed to maintain the same level of thermal comfort.

SHE have, to some extent, quantified this effect, in that standard practice is for the teleswitch operator to deduct 2°C from the temperature forecasts when the wind speed exceeds 5.66 m/s. Its inclusion within the forecasting model will help the model make more accurate inferences about the teleswitch regime leading to better predictions.

6.5.2.1 Correlation analysis on weather variables

Teleswitched demand-side management used by SHE concentrates the weather-demand sensitivity over specific half-hours. Correlation analysis is used to investigate the relationship of temperature and wind-speed with demand across the day to give an indication of the variation of the relationships across the day.

Figure 6.7 shows the correlation between the demand and the temperature and wind speed over 24 hours. The correlation was performed over 3.5 years data, from January 1995 to March 1999.

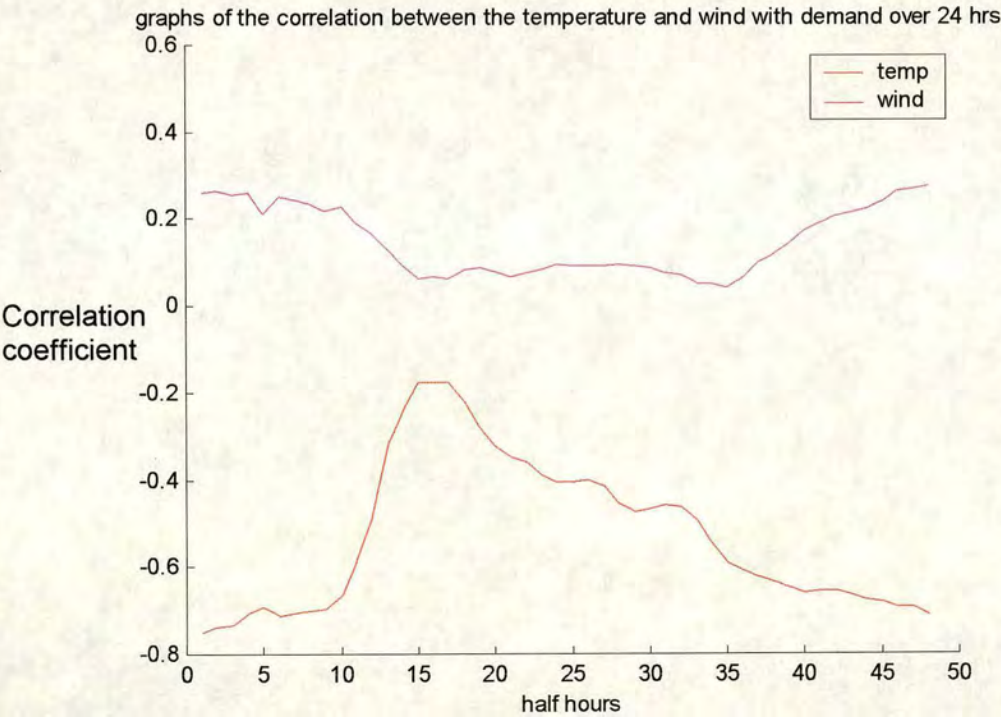


Figure 6.7. Correlation between the temperature and wind speed with demand over 24 hours

From the graph, there is a small positive correlation between wind-speed and demand during the nights, the correlation coefficient increases from zero at 18:00 hours, peaking at 0.3 around

midnight, then decreasing to zero by 0700 hours. This coincides with the heating blocks from the customers on the economy tariff and the evening and overnight THTC and RHT load blocks.

The temperature-demand correlation exhibits similar time variation to the wind speed. The correlation is always negative, confirming the inverse relationship between temperature and demand and is much stronger than that with wind speed. The negative correlation is especially strong overnight, when majority of the weather-sensitive load is switched ON. The temperature-demand correlation is lowest during the morning peak, between 0700 and 0900, suggesting that demand over this period is least temperature and wind-speed sensitive.

Due to the relationship between the weather and demand, it was decided to include half-hourly temperatures and wind speeds within the forecast models. The temperature has a lag effect on demand because it takes a period of time for building surfaces to adjust to external temperature. To account for this, the average temperature from the previous day is included as an input variable.

It should be noted that the correlation analysis captures only the linear component of the weather-demand relationship. As this relationship is very complex, the analysis does not account for the nonlinear component.

6.5.3 Correlation analysis on week day effect

Analysis of historical data, in Figure 6.5, suggested a day of week effect. It follows that demand patterns on similar days from the previous week might contain predictive information. This hypothesis is investigated in this section.

Over a period of one year, the dataset was divided into up according to the day of the week, yielding seven groups. For example, demands on all Sundays would be in the Sunday group and so on. These day-of-the week groups were correlated with each other. For simplicity, only the graph of the correlation between Wednesday's demand group and the demand groups on the other days of the week is illustrated in figure 6.8: correlations between the other demand groups showed similar patterns.

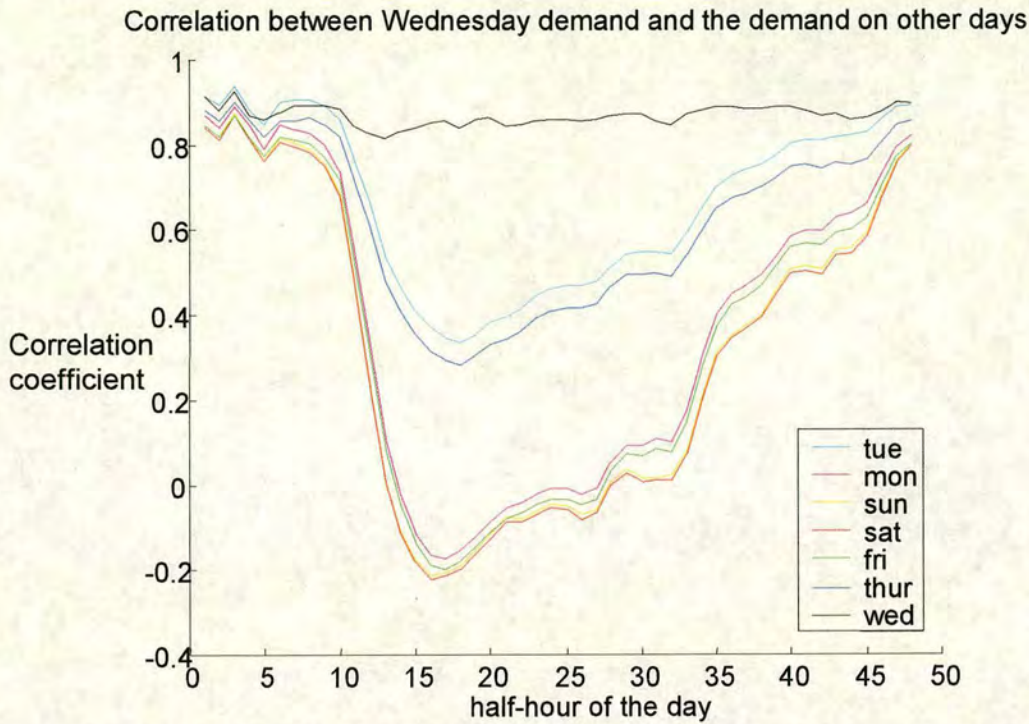


Figure 6.8. Correlation between Wednesdays demand and demand on other days of the week.

Wednesdays show a high positive correlation with themselves over all the half-hours of the day, suggesting a very strong relationship between demand patterns on Wednesdays. The correlation with other days gets less the further away from Wednesdays, with Tuesdays and Thursdays reasonably well correlated but Saturdays and Sundays poorly correlates.

Additionally the correlation coefficient between Wednesdays demand and the other days of the week vary considerably across the day, with strong positive correlation from midnight till 0600. This is probably because the demand over this period is dominated by the domestic overnight heating load, a load that is largely independent of the day of the week.

During the day, from 0700 to 1700, the relationship between Wednesdays demand and demands on Mondays, Fridays, Saturdays and Sundays is generally uncorrelated.

In summary, the above analysis highlighted a strong day of the week effect; demand on a specified day from a given week will contain information about demand on the same day, a week later. In other words, the previous week's demand is a good predictor of the following week's demand and for this reason, it was decided that, for each half-hour, the half-hourly demand from the previous week would be included as one of the explanatory variable.

Demand from the previous day was considered, but because the correlations varied considerably across the day, the predictive power of the previous day's demand was limited.

6.5.4 Correlation analysis on the effect of previous half-hourly demand

It is important to determine whether demand from a given half-hour has any relation with demand in a later half-hour and to check if previous demand information might contain information about the future demand. For example, it could contain information about random factors, such as Industrial action, which would affect demand in future half-hours.

Correlation analysis was performed between each half-hour of the day and each of the preceding 48 half hours to determine the relevant previous half-hours to be included as explanatory variables in the forecast model. The results are shown in Figure 6.9. For clarity, correlation coefficients are shown on six graphs, each representing 8 half-hours.

The correlation coefficient between the half-hour 0230 and the other half-hours of the day is very low (Figure 6.9(a)). The reason for the sudden demand reduction over this half-hour is not known but does not correspond to the switch-OFF times of any of the heating loads.

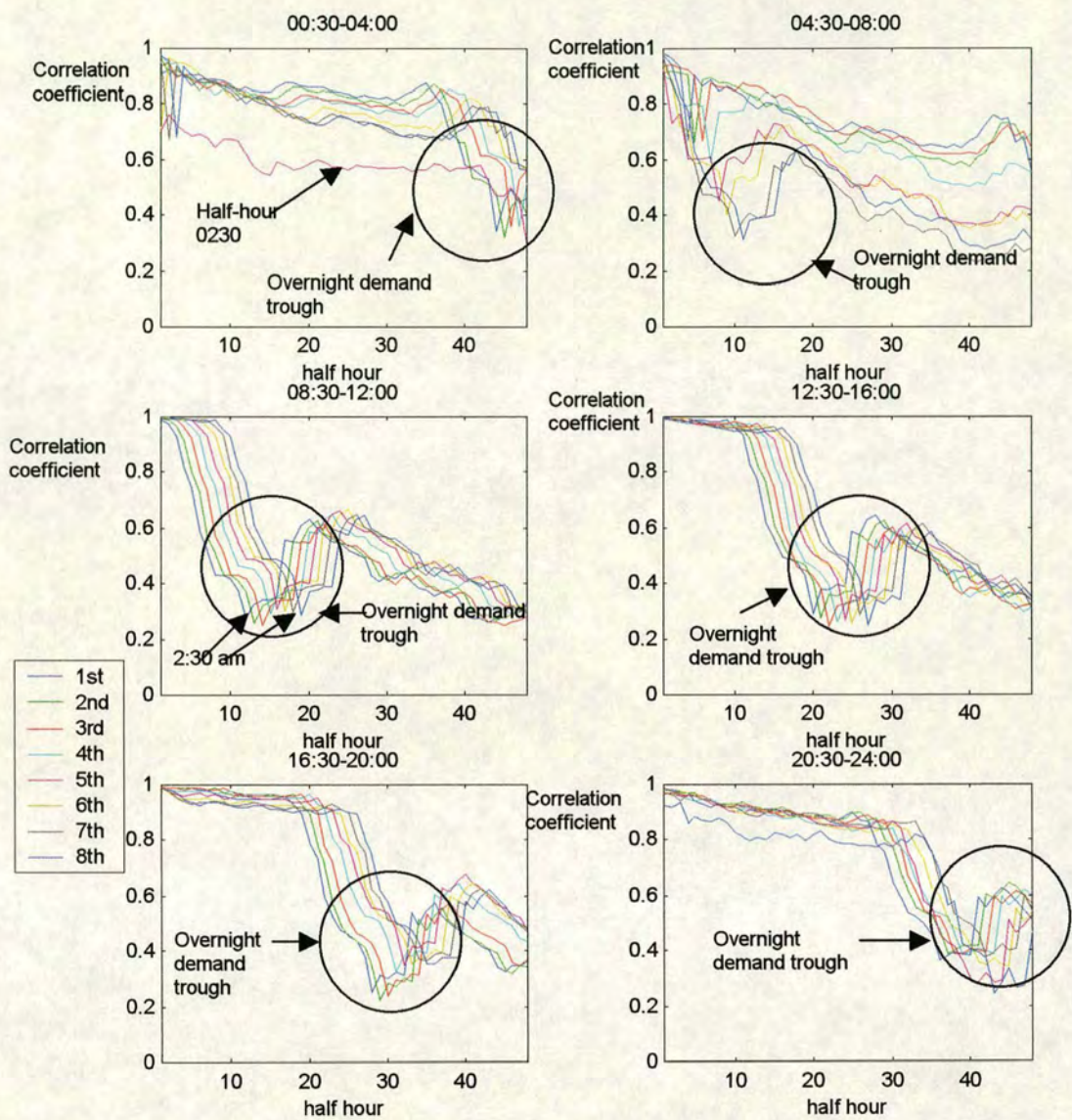


Figure 6.9 a,b,c,d,e,f. Correlations of each half-hour with each preceding 48 half-hour
the x-axis are the preceding half-hours and the y-axis the correlation coefficient

It can be seen that the correlation reduces as the half-hour considered moves in time from the reference half-hour.

This confirms the suggestion that the half-hours immediately preceding the half-hour in question will contain relevant information about the demand. As such they will be used within the forecast model as explanatory variables. For the purpose of the model comparison, it was decided to use the four preceding half-hours, because they have the highest correlation with the demand.

NETA's balancing market requires Suppliers to submit their demand requirements at BSP level to the system operator, 4 hours before the scheduled half-hour. This will mean that, at the time of Gate Closure, the forecaster will have to provide demand predictions 4 hours in advance and that inputs to the model will indeed represent forecasts. For simplicity, actual historic demand from the 4 preceding half-hours after the scheduled period will be used for the purpose of the comparisons. These will be replaced by forecast inputs for the optimisation of the preferred method.

6.5.5 Model Inputs

The analysis in the previous section indicates that the most influential explanatory variables to predict the demand are

- Forecasting will be done according to the half-hour of the day; therefore day of the week indicators will be used.
- Half-hourly temperature and wind-speed data will be used to represent the relevant weather conditions during the forecast period
- Mean temperature from the previous day will represent the lagged temperature effect
- The previous demand variables will be demand from the four immediately preceding half-hours

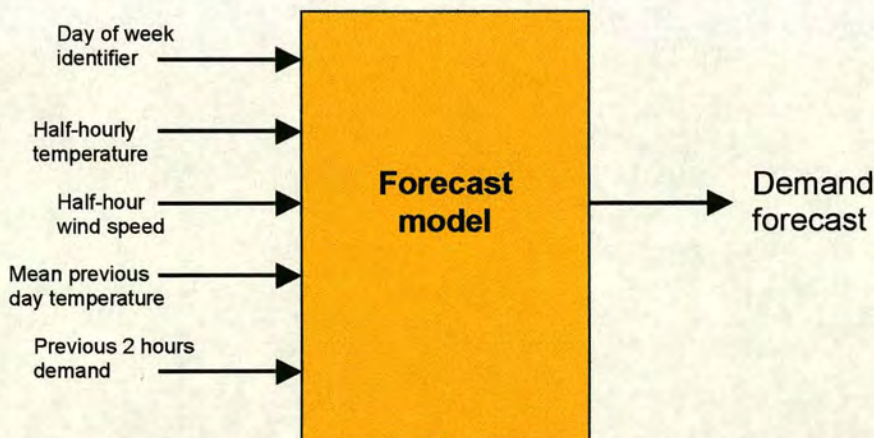


Figure 6.10. Schematic diagram of the forecast model inputs

The schematic diagram of the forecast model is shown in Figure 6.10. Though the previous week's demand on the specified half-hour is a relevant predictor, it was omitted from the first stage of the analysis to as it was felt that it was the weakest of the previous demand predictors.

6.6 Summary

In Summary, the data requirements are essential to a correctly specified model. It was decided to select a BSP in the Aberdeenshire region because firstly, it supplied a predominantly domestic load, secondly, an uninterrupted data set from January 1995 to March 1999 was available and thirdly, there was a weather station within its area as such the weather data was relevant to the demand. The models would be trained with historic data from a 3.5-year period, deemed to be optimal.

SHE use demand-side management extensively, notably teleswitching, to control the domestic heating demand. This adds extra nonlinearity into the demand, which complicated and model's ability to infer robust relationships from which to perform accurate forecasts.

The two most commonly used forecasting methods, parallel and linear forecasting were compared. It was decided to use parallel forecasting as the preferred configuration because it is assumed that the time of day variation was greater than the day of week variation.

Correlation analysis and visual inspection of data were used to determine the most appropriate inputs: these were found to be the preceding 2 hours demand, previous week's half-hourly demand, half-hourly temperature and wind-speed and mean temperature from the previous day.

7. Model construction and comparison

7.1 Overview

The requirement for accurate demand forecasting imposed by NETA has led to the present examination of the most effective short-term forecasting methods to meet Suppliers requirements. The methods selected for comparative analysis were Multiple Linear Regressions and the Artificial Neural Networks. The data requirements for the forecasts models were considered in the previous chapter and the best inputs for predicting demand at the Dyce BSP were determined from statistical analysis.

In this chapter the model construction is discussed and predictions made. Initially the forecast evaluation method used in the assessment is described followed by an outline of the issues considered in the construction of a correctly specified Multiple Linear Regression forecasting model. Details are given of the model construction and validation prior to discussing the Artificial Neural Network data pre-processing, training and construction. The most commonly used methodology within the ESI for performing naïve forecasts is presented and is used to illustrate the superiority of mathematical forecasting models over intuitive methods. Finally the results of the comparisons are presented and discussed: the chosen model is optimised in the next chapter.

7.2 Forecast evaluation

There are numerous methods for validating a forecast model. The majority of these involve the comparison of ex-post forecasts and historical data. Formal test procedures use either a Mean Absolute Percentage Error (MAPE) or a Mean Squared Error (MSE). The advantage is that deviations of a given magnitude have the same error, independent of the inputs and errors on other outputs. They give a standard measure of error, enabling different forecasts over the same time period to be compared.

$$MSE = \frac{1}{N_T} \sum_{N_T} \sum_l (L_{actual} - L_{forecast})^2 \quad \text{Equation 7.1}$$

N_T = Number of data points

L_{actual} = Actual demand at each data point

$L_{forecast}$ = Forecast demand at each data point

MSE is used during the ANN training procedure to select the optimally trained model. The most serious disadvantage of MSE measurements is that they fail to distinguish between minor and serious errors and consequently larger errors weight the MSE calculations.

The Mean Absolute Percentage Error, (MAPE) technique eliminates this bias. It is identical to MSE except the individual errors are not squared, rather their absolute values are taken. MAPE is used for the model comparison and the choice of the most appropriate ANN configuration in chapter 8, because the errors are represented as absolute percentage values. MAPE is outlined below.

Firstly, the Absolute Percentage Error (APE) is computed as

$$APE = \frac{|L_{actual} - L_{forecast}|}{L_{actual}} \times 100$$

The MAPE is derived from APE as:

$$MAPE = \frac{1}{N_T} \sum_{N_T} APE \quad \text{Equation 7.2}$$

7.3 Multiple Linear Regression

The MLR method represents the relationship between the demand and its explanatory variables as a straight line for each half-hour. It was chosen for the comparison because of its simplicity and wide applicability within the ESI.

7.3.1 Modelling issues

Before commencing on the specification of the regression model, it was necessary to investigate some of the factors and techniques that need to be considered in the construction of a correctly specified model. These factors are discussed in the following sub-sections

7.3.1.1 Dummy variables

The explanatory variables for demand fall into two categories; quantitative values such as temperature, wind speed and previous demand, and qualitative or symbolic values such as day-of-week. Qualitative inputs indicate the presence or absence of an 'effect' and cannot be defined quantitatively. They are also known as categorical inputs.

One method of ‘quantifying’ these qualitative inputs is to construct artificial variables that take on values of 1 or 0, 0 indicating the absence of the attribute and 1 indicating its presence. These are known as dummy explanatory variables, and an example of an MLR model with dummy variables is presented in equation 7.3

$$y_t = \beta_0 + \beta_1 x_t + D_1 x_t + D_2 x_t + \varepsilon \quad \text{Equation 7.3}$$

y_t	=	Electrical demand at time t
x_t	=	Explanatory variable correlated with demand
ε	=	A random error term with zero mean and constant variance
β_0, \dots, β_1	=	Regression coefficients
D_1, \dots, D_2	=	Dummy variables

An important consideration to make when using dummy variables is the dummy variable trap. This occurs when there is a perfect linear relationship between the dummy variables. For example, if the dummies in equation 7.3 ‘quantify’ two qualitative variables. In that case $D_1 = (1 - D_2)$ or $D_2 = (1 - D_1)$ suggesting that D_1 and D_2 have a perfect linear relationship. Consequentially, it would not be possible to obtain unique estimates of these qualitative parameters in such instances. A general rule of thumb when using dummy variables is to use $(m-1)$ dummy variables where m is the number of qualitative variables. In theory, dummy variables can alter either the intercept term or the slope of the regression line, or both.

7.3.1.2 Multicollinearity

Classical multiple regression models use the method of Ordinary Least Squares (OLS) to estimate regression coefficients. One of the key assumptions within OLS estimations is that there should be no exact linear relationship between two explanatory variables. Multicollinearity exists if this assumption is violated, for example, the dummy trap example. In such cases, unique estimates of all parameters cannot be obtained and statistical inferences cannot be drawn about each parameter’s influence on the regression line.

In practice, there is usually a varying degree of correlation between the explanatory variables. If this correlation is very high, it is known as imperfect multicollinearity. The object of a forecast is predictive and not a reliable estimation of the individual contribution of each parameter to the regression line. The inclusion of a specific explanatory variable could introduce imperfect multicollinearity, however if it increases the predictive capability of the model, then its inclusion is desirable.

7.3.1.3 Heteroscedasticity

Another assumption of the OLS method is homoscedasticity, that is, the variance of the error term ε is constant and has zero mean. Heteroscedasticity occurs when there is uneven or unequal variance among the error terms of the various observations. This could bias the confidence interval measurements thus affecting hypothesis tests based on the F and t statistics.

Research has shown that heteroscedasticity is usually found in cross-sectional data and not in time-series data. In cross sectional data, the members of a population are observed at a given point in time, unequal variance usually results because some members of the population tend to have larger variation in magnitude. With time series data, variables tend to be of similar orders of magnitude because data is collected for the same entity over a period of time. Short-term demand forecasts fall under the time-series banner, as such heteroscedasticity is not a particular problem.

7.3.1.4 Autocorrelation

The final assumption for the OLS method states that no correlation exists between the error terms. This means that the disturbance term relating to any observation is not related to or influenced by the error term relating to any other observation. For example, the error on the OLS estimation of Monday's demand should not necessarily be related to the error on Tuesday's demand nor the demand on the following Monday. Auto correlation exists when, for example, lower demand caused by high temperature on Monday affects demand for the remainder of the week.

Just as heteroscedasticity is generally associated with cross-sectional data, autocorrelation is usually associated with time series data. The consequences of autocorrelation are very similar to heteroscedasticity. Standard formulas underestimate the true variance and standard errors, thereby inflating the t -statistic. This gives the impression that a particular coefficient is statistically different from zero, whereas in fact that might not be the case. As a consequence the computation of R^2 , a measure of the overall goodness-of-fit of the regression line, may be unreliable. This will affect the efficiency of the computed variance and standard error of the forecast.

For the MLR forecast model, it is very important to check for autocorrelation to determine the validity of the statistics associated with the regression line. The simplest approach is a visual examination of the OLS residuals against time.

7.3.2 Model construction

The previous section presented the issues considered when building a robust Multiple Linear Regression forecasting model using, most notably Heteroscedasticity, Autocorrelation, Multicollinearity and dummy variables. This section considers the construction of the model and assesses its validity.

To summarise the discussion in the previous chapter, a parallel forecasting configuration will be used because of the assumption that the time of day variations at the Dyce BSP are greater than the day of week variations. This forecasting configuration will require 48 identical models, each assigned to learn and forecast the demand on a specified half-hour of the day. The explanatory variables for the demand are half-hourly temperature and wind-speed, mean temperature from the previous day, previous 2 hours demand and day-of week identifiers.

Dummy variables are used to encode the symbolic day-of-week inputs. Since there are seven days of the week, six dummy variables would be used to avoid the dummy trap. The MLR model is illustrated below in equation 7.4

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 \bar{x} + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} + \rho_4 y_{t-4} + D_1 x_t + D_2 x_t + D_3 x_t + D_4 x_t + D_5 x_t + D_6 x_t + \beta_3 z_t + \varepsilon_t$$

Equation 7.4

- y_t = Independent variable, future demand
- β_0 = Intercept term or base demand
- β_1 = Regression coefficient for the half-hourly temperature
- β_2 = Regression coefficient for the previous day's mean temperature
- β_3 = Regression coefficient for half-hourly wind speed
- x_t = Temperature during the specified half-hour
- \bar{x} = Previous 24 hours mean temperature
- $y_{t-1} \dots y_{t-4}$ = Half-hourly demand from previous 2 hours
- z_t = Wind speed during the specified half-hour
- $\rho_1 \dots \rho_4$ = Regression coefficients for the half-hourly demand over the previous 2 hours.
- $D_1 \dots D_6$ = Dummy variables for the day of the week
- ε_t = Error term

MLR models have the advantage of transparency because the regression coefficients, which give an indication of the contribution of the explanatory variables to the model, are discretely expressed. β_0 the intercept term, shows the contribution of the base demand, $\beta_1 \dots \beta_3$ the temperature and wind speed variables, $D_1 \dots D_6$ the day of the week effects and $\rho_1 \dots \rho_4$, the contribution of the previous demand.

The 3.5-year data set was divided into a training and test data set. Equation 7.4 was programmed and a training data set containing the explanatory variables as inputs and the corresponding demand as output over a three-year period was applied to the model to derive the regression coefficients. The MLR equation for the first half-hourly period of the day is shown below.

$$y_t = 4.9553 - 0.0252x_t + 0.2545\bar{x} + 0.2423y_{t-1} - 0.0692y_{t-2} + 0.4271y_{t-3} - 0.1768y_{t-4} \\ - 0.0266x_t - 0.00726x_t - 0.0347x_t - 0.0169x_t - 0.0183x_t - 0.0069x_t + 0.0021z_t$$

To perform a prediction over this half-hour, the explanatory variables are used in the equation above.

7.3.3 Model validation

Before the model is used to predict the demand, it has to be validated to ensure that it has been correctly specified. The model validation is discussed in this section starting with the test statistics.

7.3.3.1 Test statistics

The MLR represents the demand relationship as a straight-line. It is important to determine how well the regression line models this relationship, in other words its 'goodness of fit'. This is done by measuring the R^2 statistic, also known as the coefficient of multiple correlation. The R^2 statistic indicates the level of influence of the explanatory variables on the dependent variable.

R^2 is always a positive number between 0 and 1: a value of 1 means that all the variation in the regression line is explained by the explanatory variables while very high values suggest that the regression model is correctly specified. For the model derived in the previous section the R^2 measure is 0.9175, indicating that the explanatory variables represent 91.75% of the variation in demand

However, the R^2 statistic does not show the influence of individual explanatory variables on demand. To determine whether an individual explanatory variable is significant, the estimated standard deviation from the variance-covariance matrix is observed. As a rule of thumb, an explanatory variable is deleted from the model if it is less than 2 standard deviations away from zero because it has a negligible effect on demand.

The variance-covariance matrix is computed as

$$y_t = w_t' \delta + \varepsilon_t$$

$$\begin{aligned}\hat{\delta} &= (w'w)^{-1} w'y \\ \text{var}(\hat{\delta}) &= \hat{\sigma}^2 (w'w)^{-1} \\ \hat{\sigma}^2 &= \frac{1}{n-k} \sum_{i=1}^n e_i^2 \quad e_i = y_i - w' \hat{\delta} \quad \text{Equation 7.5}\end{aligned}$$

y_t = vector of demand variables

w' = matrix of regression coefficients

$w' \hat{\delta}$ = vector of least squared estimators

$\hat{\sigma}^2$ = standard deviation of the variance covariance matrix

The next step in the model validation was to determine whether each of the estimated regressors are relevant within the forecast model from the variance-covariance matrix. The estimated standard deviations of the explanatory variables were observed and all the regressors were more than 2 standard deviations away from zero. This means that the chosen explanatory variables are all relevant within the model.

7.3.3.2 Test for autocorrelation

Short-term demand forecasts were made from time-series data and could therefore be susceptible to autocorrelation. In other words the error on one observation might be related to, or influenced by the error on another observation. Autocorrelation would cause estimates of R^2 to be unreliable thus the validity of the regression's test statistics would be compromised.

A visual inspection of the regression's residuals was used as a preliminary test for autocorrelation and the residuals from the derivation of the MLR in period 1 were computed over the entire training period, 1095 days, see Figure 7.1.

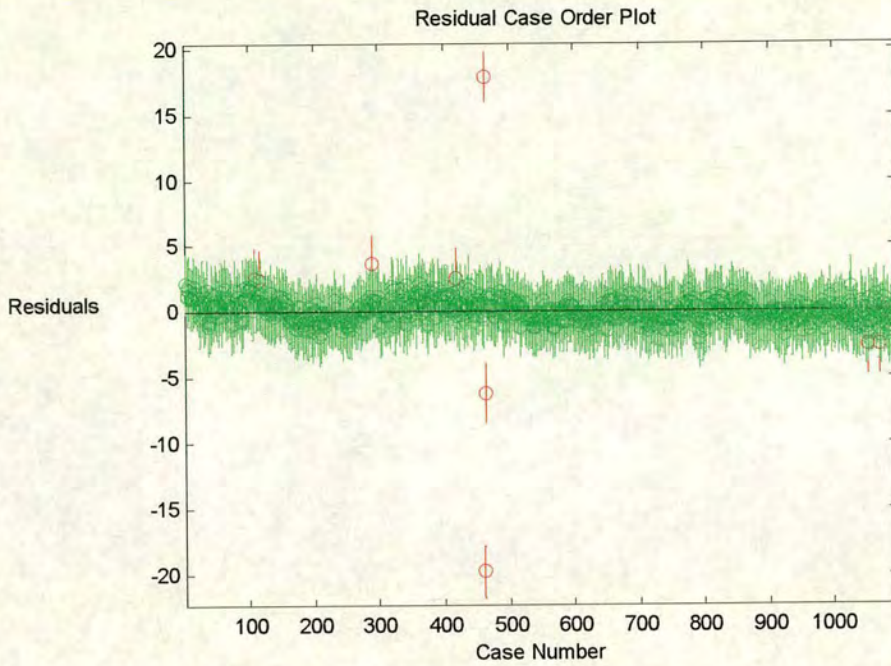


Figure 7.1. Regression residual plot over the training period

Visual inspection shows slight periodicity during the first year, from day 1 to day 365 beyond which the residuals are evenly distributed around zero. This suggests an initial small degree of autocorrelation that disappears over time. The amount of autocorrelation present in the training data set was considered negligible and the R^2 estimate of 91.75% was taken as a reliable measure of the goodness of fit. These tests suggest that the model is correctly specified.

The analysis detailed in this section is relevant to the forecasting model for period 1 only and similar validation had to be performed for the other 47 half-hourly models.

7.4 Artificial Neural Network

A brief outline was given in Chapter 5, of the basic structure of an Artificial Neural Network (ANN). Artificial Neural Networks were selected for the comparison because they are the most popular Artificial Intelligence method that may be used for short-term forecasting and are the most powerful tools for modelling multivariate 'cause and effect' relationships.

This section discusses the data preparation, learning algorithm, training and the final architecture of the ANN.

7.4.1 ANN data preparation

Numerical conditioning is one of the most fundamental and important concepts in neural network implementation because it affects the speed and accuracy of the learning algorithm. For this reason the input data is pre-processed to encode the qualitative day-of-week inputs and to present the inputs to the network so as to maximise its inference potential.

7.4.1.1 Encoding categorical information

With MLR models, dummy variables were used to encode qualitative or categorical day of the week information. Several approaches exist to do this and at the time of writing, there was none universally acknowledged as the best.

One approach assigns numbers 1 to 7 for each day of the week. This method's major advantage is that it requires only one input to represent the qualitative information. Its disadvantage is that the ANN will attach higher weighting to the days with higher numbers. For example, if Saturday is labelled day 7, and Sunday, day 1, the network might give undue weighting to Saturday's demand pattern although it is no more valid than Sunday's. This method was discarded to ensure that encoding of categorical information did not introduce any unwanted meaning.

Another family uses binary codes to represent the categorical information. The simplest approach would be to use one bit per day. Though this method is effective, it requires a 7-bit code corresponding to 7 designated input neurons. This approach was rejected because it uses an excessive number of inputs. At the opposite extreme, the optimal binary coding method, as shown in Table 7.1, uses 3 bits to encode all seven categorical inputs.

Day of week	Binary Codes		
Sunday	0	0	0
Monday	0	0	1
Tuesday	0	1	0
Wednesday	0	1	1
Thursday	1	0	0
Friday	1	0	1
Saturday	1	1	0

Table 7.1. Optimal binary encoding for categorical inputs

Though this method gives a binary representation of the data using the minimal number of inputs, it will not represent the optimal solution⁹¹. This is because Monday, Tuesday and Thursday identifiers switch one-bit, whereas Sunday switches none, whilst the others switch two bits. The

unequal number of switched input neurons could potentially favour the demand patterns on certain days.

The optimal solution is to combine both binary methods using a code with minimum inputs that switches the same number of input neurons for each day of the week. In other words, the lowest bit code with the same number of ones for each categorical input. For the 7-category day of the week case, a 5-bit code with 2 of the inputs nodes always switched is the most effective. This is shown in Table 7.2, there are 3 spare codes. This method should eliminate all possible bias towards any of the day of week categories.

Day of week	Binary Codes				
Sunday	0	0	0	1	1
Monday	0	0	1	0	1
Tuesday	0	0	1	1	0
Wednesday	0	1	0	0	1
Thursday	0	1	0	1	0
Friday	0	1	1	0	0
Saturday	1	0	0	0	1
Spare codes	1	0	0	1	0
	1	0	1	0	0
	1	1	0	0	0

Table 7.2. 5-bit day of week code

7.4.1.2 Input data pre-processing

Since neural networks can perform arbitrary non-linear functional mapping between sets of variables, a neural network can, in principle map raw data directly onto required final output values. Nevertheless, in most practical applications the choice of pre-processing will be one of the most significant factors in determining the performance of the final network.

One of the most common forms of pre-processing requires simple linear rescaling of the input variables⁹². This is particularly useful if the explanatory variables have typical values that differ significantly in magnitude from each other. For example, temperature, wind speed and previous demand are each expressed in different units and also have significant variations in magnitude. In such cases the typical sizes of the inputs may not reflect their relative importance in the determination of the required outputs. Applying a linear transformation to each input independently can solve this problem by equalising all the input values to fit a similar range.

One simple transformation is normalisation based on the mean and standard deviation. For each explanatory variable x_i , the mean \bar{x} and variance σ^2 is calculated using.

$$\overline{x_i} = \frac{1}{N} \sum_{n=1}^N x_i^n$$

$$\sigma_i^2 = \frac{1}{N-1} \sum_{n=1}^N (x_i^n - \overline{x_i})^2$$

Equation 7.6

where $n = 1, \dots, N$ labels the patterns numbers. The rescaled values for each explanatory variable are calculated using the equation given below.

$$\tilde{x}_i^n = \frac{x_i^n - \overline{x_i}}{\sigma_i}$$

Equation 7.7

The transformed values given by \tilde{x}_i^n have zero mean and unit standard deviation over the transformed data set. This removes all effects of offset and measurement scale.

Table 7.3a and 7.3b show a cross-section of the data file before and after pre-processing. They present the input-output patterns for period 6 on three separate days. Columns 1-12 contain the input data and are presented to the input of the neural network whilst column 13 contains the output data and is presented to the network's output. Table 7.3b shows the data after it has been pre-processed; the output data is not pre-processed.

5-bit day of the week categories					Mean temp °C	Temp °C	Wind Knots	Previous 2 hour demand MW				Demand
0	1	0	1	0	13.4	12.6	6	18	18	20	17	17
1	0	0	0	1	12.5	12.6	5	18	18	19	16	16
0	0	1	0	1	11.8	13.2	10	15	16	17	15	15

Table 7.3a. Neural Network data before pre-processing

5-bit day of week category					Mean T °C	T °C	Wind Knots	Previous 2 hour demand				Demand
-0.4	1.14	-0.8	1.14	-1.14	1.46	1.35	-0.4	-0.8	-0.99	-0.7	-0.99	17
2.5	-0.87	-0.8	-0.87	0.87	1.23	1.35	-0.6	-0.8	-0.99	-0.9	-1.2	16
-0.4	-0.87	1.2	-0.87	0.87	1.1	1.48	-0.17	-1.8	-1.65	-1.6	-1.5	15

Table 7.3b. Neural Network data after pre-processing

In summary, the inputs to each network are shown in the schematic diagram in figure 7.2

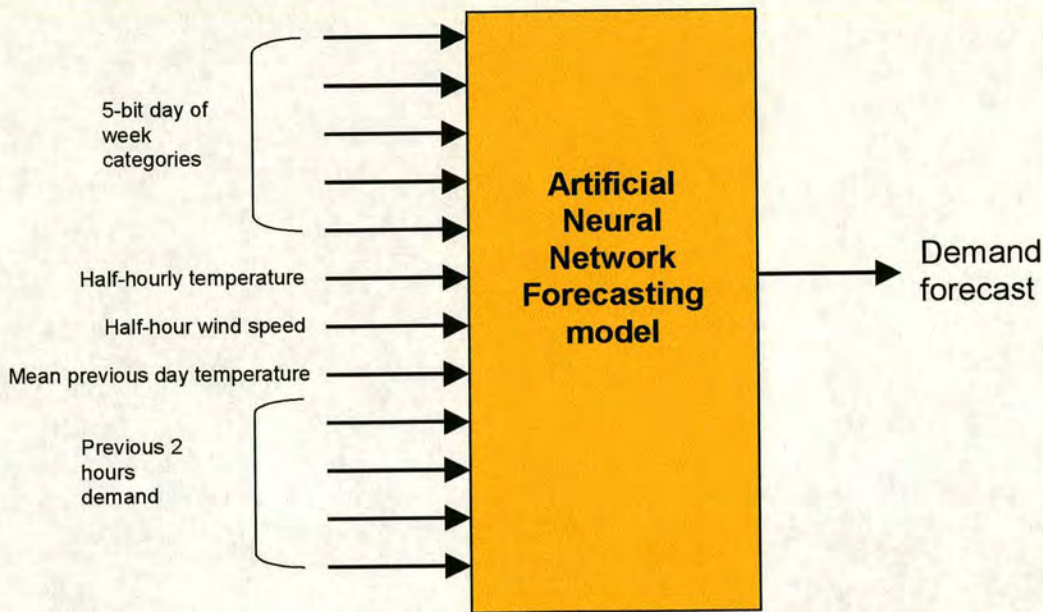


Figure 7.2. Schematic diagram of the ANN's inputs

7.4.2 Training the network

The development of the back-propagation algorithm in the mid-1980s represented a landmark in neural computations because it provided an efficient method for training multi-layered networks⁹³. Once trained these networks have the ability to 'learn' from examples and generalise beyond the training data set.

As a training algorithm, the purpose of back-propagation is to adjust the network weights so the network produces the desired output in response to the input pattern from a predetermined set of training patterns. It is a supervised algorithm, for every input pattern, there is an externally specified 'correct' output which acts as a target for the network to imitate. Training is done 'off-line'; the network is used afterwards in normal operation, no further learning occurs after the training phase.

To train a network, a set of input patterns and corresponding desired outputs is used along with an error function to measure the differences between the neural network's outputs and the desired values. The most commonly used error function is the Mean squared Error (MSE). Its major advantage is the ease of differentiability and the fact that the error calculation depends on its magnitude only. As such a deviation of a given magnitude has the same error independent of the input pattern and errors on other outputs.

The basic training steps of the back-propagation algorithm are summarised below.

- A training input pattern is presented to the network, and propagated through to obtain outputs.
- Compare the outputs with the desired values and calculate the error.
- Calculate the derivatives $\partial E / \partial w_{ij}$ of the error E with respect to the weights w_{ij}
- Adjust the weights to minimise the error
- Repeat until the error is acceptably small or the maximum number of epochs specified at the beginning has been exceeded.

A flow diagram of the training process is illustrated in Figure 7.3. One such pass through the network resulting in updating of the network's weights is called an epoch.

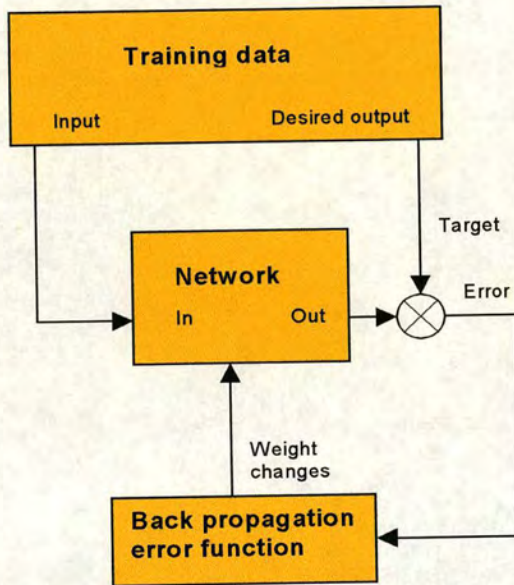


Figure 7.3. Flow diagram of the training process using a back-propagation algorithm

In practical applications of the back-propagation algorithm, learning results after many epochs. There are two types of learning; Pattern and Batch mode training. For the purpose of this discussion, Pattern mode training is favoured because the start point or seed of the input-output patterns is chosen at random as such the network is less likely to get trapped in a local minimum.

Once adequately trained, the network should be capable of performing forecast from an input data set it has never 'seen'. This is known as generalisation. The learning process could be viewed as a curve-fitting problem, where the network performs a non-linear interpolation of the input data.

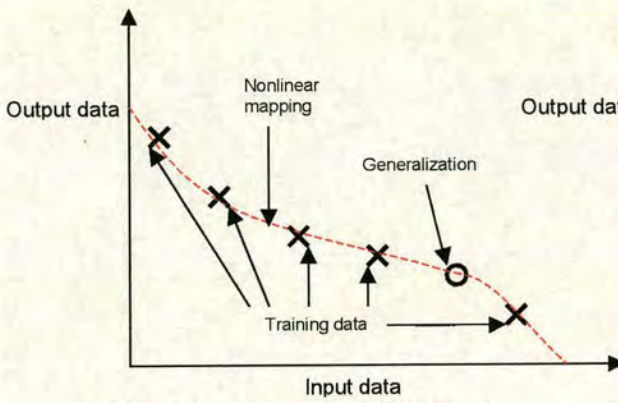


Figure 7.4a. Properly fitted data

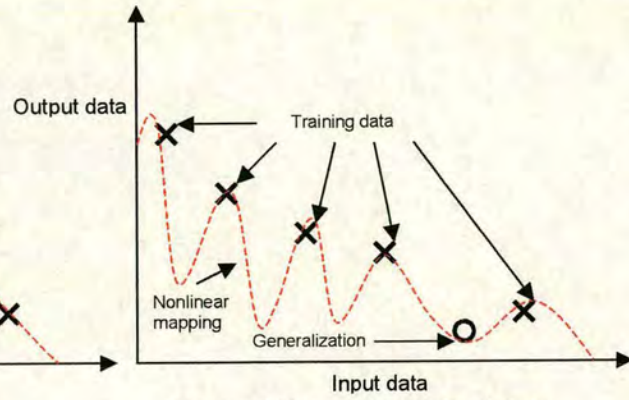


Figure 7.4b Over fitted data

A well trained neural-network will have good generalisation and is capable of producing correct input-output mapping even when the input differs from the training data. A properly trained network with good generalisation is shown in figure 7.4a. However, when a neural network learns too many specific input-output relations, the network may memorize the training data and is less able to generalise between slightly different input-output patterns. This is known as over training and is illustrated in figure 7.4b. In this example the neural network has memorised the data, when forecasting, the neural network acts as a lookup table and will perform poorly on data it has never seen. It is therefore very important to prevent over training.

The reverse is under training, which occurs when the network does not have sufficient input-output pairs to infer effective relationships. Under training is avoided by presenting the network with a large number of input-output parings.

7.4.3 Designing a well trained model: Cross validation

The next question is how to optimise the network so as to avoid over training and achieve good generalisation. One of the simplest approaches is cross-validation. The data set is randomly partitioned into a training set and a test set. The training set is further partitioned into two subsets

- A subset used to train the network known as the training data set.
- A subset used to evaluate the performance of the model during training, known as the validation data set.

Whilst training, the Mean-Squared Error (MSE) of the network decreases as the number of epochs increases. To assess the models generalisation performance, the network is tested after each epoch on data it has 'never seen', the validation data set. The validation error also decreases at the start of training. Training is continued until the validation error starts to rise. With this method the performance of the network is monitored during training until the best

performance is attained, known as the optimal training point. Cross-validation avoids over training. An example of cross-validation is shown in figure 7.5

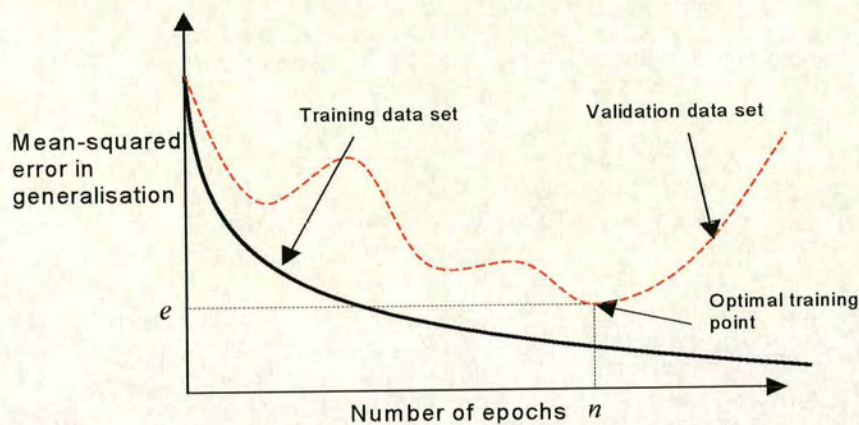


Figure 7.5. Cross-validating the models during training

During training the validation error occasionally decreases past its initial minimum. For this reason training is continued well beyond the optimal training point. After this the model is then retrained starting from the seed of the previous run until the optimal training point. The model's generalisation performance is finally tested on the test data set. A flow diagram of the training process is shown in figure 7.6

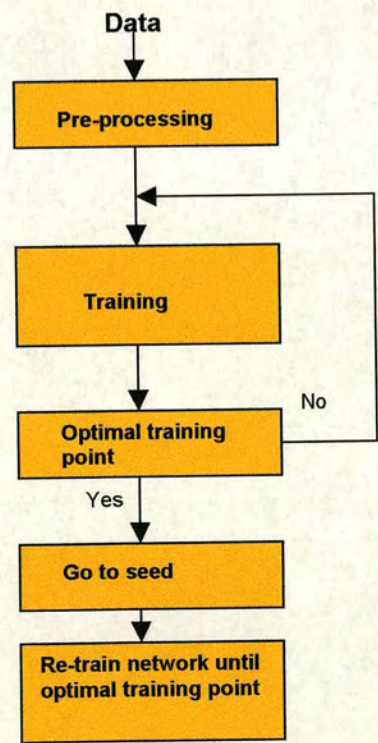


Figure 7.6. Flow diagram of the training process

With the back-propagation algorithm training starts from a random initialisation point in the training data set. As such, the optimal training point for any two networks will be different because they have different start points. Thus networks trained on the same data set will have slightly different accuracies. It is common practice to train many different candidate networks and select the best, based on their performance on the independent test data set. A reasonable number of candidate networks are about 10. This is known as the hold-out method.

7.4.4 Network architecture

The network architecture involves selecting the optimal number of hidden layers and the number of neurons within each hidden layer. In multiple-layer feed-forward networks, the number of hidden neurons can make the difference between a well-specified network. There are no hard-and-fast formulas for doing this, simply general suggestions.

Generally, there are no real benefits for using more than one hidden layer because ANN training slows down dramatically due to the following effects.

- The additional layer through which errors must propagate makes the network unstable.
- The number of local minima increases, and can trap the network. When this happens, training has to be restarted.

For the purpose of short-term demand forecasting it was decided to use one hidden layer.

Choosing the appropriate number of hidden neurons is extremely important. Using too few neurons will starve the network of the resources it needs to infer relationships between the explanatory variables and demand and also increases training time. An excessive number of hidden neurons could cause over fitting.

The best approach to finding the optimal number of hidden neurons is by experimentation with various sizes, starting from the least number. Five architectures were compared; 3 neuron, 5 neuron, 7 neuron, 10 neuron and 15 neurons. For each architecture, 10 networks were built for both half-hours. The networks with the lowest MSE for each architecture are presented in Table 7.4

During training there was a noticeable convergence time difference, with the 3-neuron version taking the longest. From the results and the trade-off with training time, it was decided to use the 5 hidden neuron architecture.

Architecture	Period 1	Period 25
3 neurons	0.39	0.36
5 neurons	0.21	0.24
7 neurons	0.27	0.25
10 neurons	0.25	0.24
15 neurons	0.25	0.26

Table 7.4. Comparison of various hidden neuron architecture

7.5 Naïve forecasting

For its day-ahead forecasting, NGC uses a forecasting method based on demand from the previous three days, a method very commonly used in the ESI. For the purpose of this discussion it is referred to as a naïve forecast. This method calculates present demand as the sum of the weightings of the demand from the previous 3 days. The results from a naïve forecast over the time period concerned are presented to demonstrate the superiority of both methods to the industry's average. The weighting is as follows:

$$y_t = \frac{1}{7}y_{t-3} + \frac{2}{7}y_{t-2} + \frac{4}{7}y_{t-1} \quad \text{Equation 7.8}$$

where y_t is the forecast demand.

7.6 Results and discussion

The complete forecasting model would require the construction and training of 48 models for both ANN and MLR methods, totalling 102 models. Such a large number of models would be unnecessarily excessive for the comparison. Instead it was decided to perform the comparison over selected half-hours that characterised the demand at the Dyce BSP. This would shorten the time taken without diminishing the validity of the comparison.

The chosen half-hours were period 1, 00:00-00:30 and period 25, 12:30-13:00. Period 1 was chosen because it reflected the over night demand. It contained a high proportion of the heating load and had very limited day of the week variation. Period 25 was chosen because it was one of the marginal half-hours, from the discussion in section 6.3.5, the nonlinearity across such periods presented the greatest challenge to forecasting models.

The same data set was used for both models. The training set consisted of three years of data, from January 1996 to December 1998. The forecasting capability of both models was determined over the test period, from January 1999 to March 1999.

The MAPE for both models together with the naïve forecast for both periods 1 and 25 are presented over the 90 day forecast horizon in Table 7.5.

Model	Period 1	Period 2
Naïve forecast	4.6746	17.6287
Multiple Linear Regression	2.8442	4.0101
Artificial Neural Network	1.8630	1.0314

Table 7.5. MAPE for all the models

The forecasts from both models are superior to the naïve forecast notably during period 25. This shows that a mathematical forecast model reduces Suppliers volumetric risk considerably. The ANN outperforms the MLR over both half-hours. However, there is not much difference between the predictive capabilities of both models for the first half-hour, this is because Period 1 has fewer nonlinearities. Demand over this time period is not affected much by day of week variations, its primary driver is the weather because of the large element of stable heating demand. The MLR is thus capable of making fairly accurate inferences.

The situation is rather different during the second half-hour. The ANN performs better whereas the quality of the MLR forecasts worsens. This suggests that the ANN is very capable of modelling and predicting the nonlinearity imposed by teleswitching. As the MLR is only capable of representing the relationship linearly, it is incapable of capturing the nonlinearity imposed by the teleswitch block that is switched ON for minimum forecast day ahead temperatures below 1.

Type of forecast	Period 1	Period 25
Naïve	94.82MW	498.6MW
MLR	60MW	128.5MW
ANN	38MW	29.2MW

Table 7.6 Forecast error in MW terms over 90 day period

The forecast error expressed in demand terms is shown in Table 7.6. The error is the out-of-balance quantity Suppliers using the different forecasting methods will be penalised for by the System Operator (SO). The numbers show the superiority of the ANN method over MLR and the virtues of mathematical forecast.

The forecast from both models are presented graphically in figures 7.5 and 7.6. Figure 7.5 shows the actual demand and the forecasts for period 1 whilst Figure 7.6 presents period 25 over the 90-day time horizon.

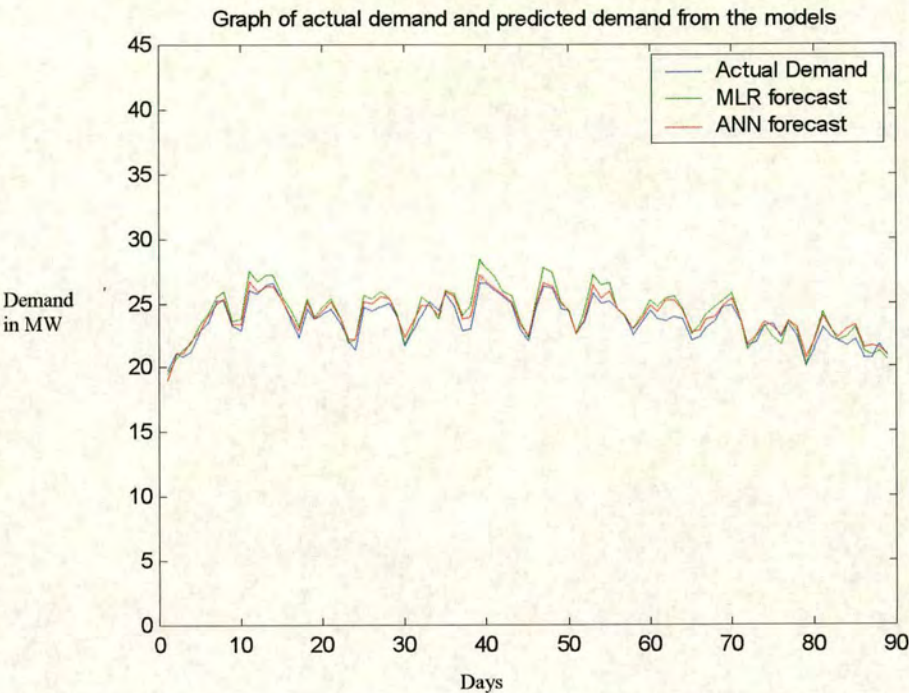


Figure 7.5. Actual demand and forecast demand during period 1 for a 90-day horizon

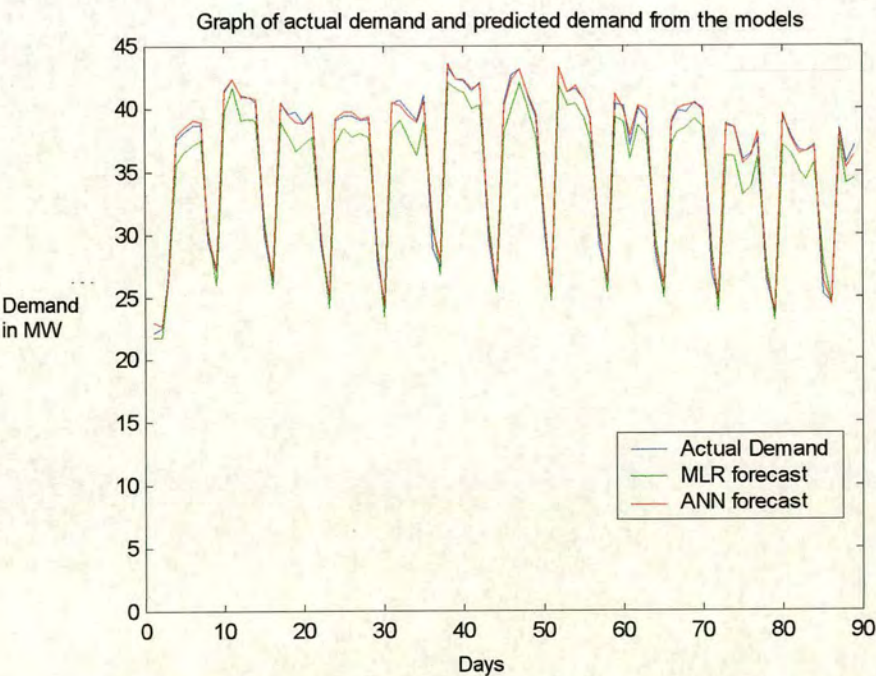


Figure 7.6. Actual demand and forecast demand during period 25 for a 90-day horizon

Both graphs show a cyclical pattern, the peaks represent the weekday demand and the troughs, the weekends. In Figure 7.5 both models occasionally over predict the demand with the MLR indicating slightly greater overpredictions.

Over period 25, in figure 7.6, the ANN prediction is much improved and closely follows the actual demand. On the other hand, the MLR's predictions deteriorate. The MLR model consistently under predicts the weekday demand. It is probable that this occurs because the MLR model is unable to represent the wide demand discrepancies between the weekday and weekend demand as such, the predicted weekday demand is reduced by a constant magnitude. The phenomenon can be partially corrected by scaling up the weekday demand. However, this was not considered appropriate, as the analysis was intended to be a straight comparison between both models.

Next, it was decided to observe the performance of both models across the temperature range they were forecasting during the 90-day period. This is to assess whether the temperature affects the predictive accuracy of either model. Table 7.6 shows the MAPE for both models for 3 temperature ranges.

Model	Period 1			Period 25		
	$T < 0^{\circ}$	$0^{\circ}\text{C} \leq T \leq 5^{\circ}\text{C}$	$T > 5^{\circ}\text{C}$	$T < 0^{\circ}$	$0^{\circ}\text{C} \leq T \leq 5^{\circ}\text{C}$	$T > 5^{\circ}\text{C}$
MLR	5.5151	3.0701	1.4519	2.2093	3.8007	4.2285
ANN	1.94987	1.9438	1.4999	1.4071	0.8754	1.0771

Table 7.7. The MAPE of both models over the forecast temperature range

From the results in the Table 7.7, the MLR makes inaccurate predictions at temperatures below 0 degrees during the first period, where temperature is the primary driver of demand because the majority of the over-night space-heating demand is active over this half-hour. Because the temperature-demand relationship from the discussion in section 5.3.4.1 follows an inverse parabolic curve, the MLR's straight-line approximation will not account for data points at the ends of the curve. The MLR represents demand at the upper temperature range better because it is closer to the mean of the training data set. On the other hand because the errors from the ANN model are consistent over the temperature range, this illustrates that the ANN is capable of representing the parabolic nature of the temperature-demand relationship.

For period 25, the ANN performs particularly well over the temperature range between 0 and 5 degrees where it was expected the most modelling complexity would exist because the teleswitch demand block is arranged to be ON for temperatures below 1 degree and OFF for temperatures

above. However, the ANN is capable of modelling this non-linearity, because predictions over this temperature range have the lowest errors.

During period 25 the MLR's errors are fairly large especially for temperatures above 5 degrees. This could be because the nonlinearities induced by teleswitching weights the regression line in favour of lower temperatures, biasing the forecast against higher temperatures. As such the magnitudes of the error increases as the temperature increases.

The ANN model is superior for forecasting demand at this BSP, adequately modelling both nonlinearities within the temperature-demand relationships and teleswitching. The straight-line approximation of the Multiple Linear Regression model is not capable of representing these nonlinearities, as such the accuracy of its predictions are compromised. However, the MLR model has the advantage of transparency over the Neural Network model, since contributions of the individual explanatory variables on the demand are easily observed.

As an experiment seasonal dummies were introduced into both models to account for the time of year effect. These did not make any difference to the forecast accuracy, because there was no seasonal change over the forecast period and the lighting peak does not occur during either periods 1 and 25.

7.7 Conclusions

This chapter was primarily concerned with the analytical comparison of the forecasting models selected in chapter 5. Following a brief discussion of the forecast evaluation method which covered several issues involved in the correct specification of a MLR forecasting model, notably autoregression and the representation of the day of week variables, a forecasting model for the first period of the day was constructed and validated.

ANN is the most popular Artificial Intelligence tool used in short-term demand forecasting and the requirement was emphasised to encode data pre-processing of categorical information and scaling of inputs.

The need to achieve a well-trained neural network to achieve optimal performance was discussed. The back-propagation training algorithm used in the network was outlined together with the optimisation methodology, cross-validation, to select the best trained network and this section concluded with a discussion on the network's architecture.

A brief outline of naïve forecasting was given followed by the equation of the most commonly used method. Naïve forecasts were used as a benchmark to assess the merit of using either of the formal forecasting approaches.

The final section of this chapter presented and discussed the results of the forecasts from both models. A 90-day forecast period was used and the comparison was made using demand over two half-hours, periods 1 and 25 that sufficiently characterised the load. From the comparisons, the ANN outperformed the MLR in all categories assessed. The ANN was capable of modelling the nonlinearities within the demand whereas the MLR attempted to model the relationship as a straight line. Both models were superior to the naïve forecast, confirming the value of proper short-term forecasting under NETA.

The final chapter discusses and analyses various optimisation methodologies to further improve the accuracy of the ANN followed by the construction of models for each of the 48 half-hours and a comparison of linear and parallel forecasting.

8. Neural network forecast optimisation

8.1 Overview

The ultimate object of forecasting is to try, as far as possible, to insulate the Supplier from unwanted out of balance costs. Artificial Neural Networks were chosen from the preceding discussion as the most appropriate method for predicting half-hourly demand at the Dyce BSP in Aberdeenshire. This chapter initially focuses on optimising the efficiency of the ANN, followed by the complete construction and analysis of the 48 half-hourly models and comparison of linear and parallel methods.

The first step in the optimisation is to compare forecast accuracy of binary encoding methodology with an alternative. In an attempt to shorten the training time whilst improving the accuracy, the length of the training window and forecast horizon were re-examined. Before building the complete model, the validity of the chosen input architecture was assessed by constructing a model for the first six hours of the day. This revealed some inherent flaws. The input variables were re-evaluated and a revised model configured, optimised and then construction for each of the 48 half-hours was carried out.

In Chapter 6, it was assumed that parallel methods were superior to linear methods at forecasting demand at this BSP. A comparison of the performance of the completed parallel method with a linear equivalent is used to illustrate this. This chapter ends with a case study using real NETA balancing mechanism prices to confirm the validity of the forecast methods developed.

8.2 Assessing an alternative encoding method

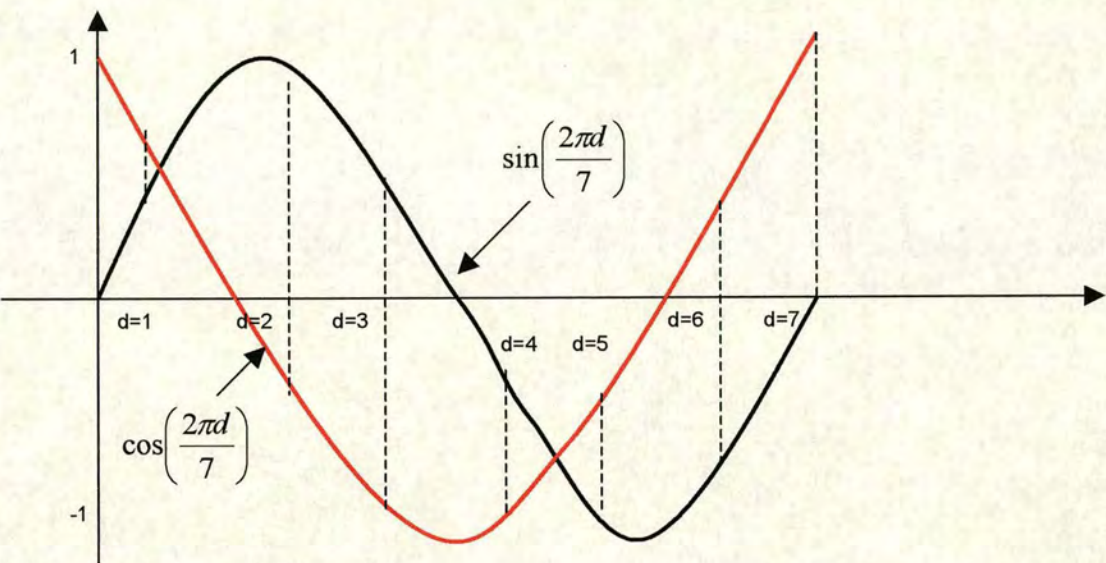
The purpose of encoding the categorical inputs was discussed in the previous chapter. The preferred method was chosen as a 5-bit binary code utilising minimal inputs with the same number of input neurons switched for each day of the week. To determine its effectiveness, it was decided to contrast this method with an alternative approach known as the sine-cosine method⁹⁴.

With the alternative approach, day of the week information is encoded as sine and cosine functions. Each day of the week is uniquely specified by a periodic measure defined by the trigonometric functions shown below.

$$\sin\left(\frac{2\pi d}{7}\right) \qquad \cos\left(\frac{2\pi d}{7}\right)$$

Equation 8.1

Both functions are used to encode the categorical information by assigning a number from 1 to 7, d , for each day of the week in both equations. Graphical representation of the day of the week



information is shown in Figure 8.1

Figure 8.1. Graphical representation of the sine-cosine function

The day of week identifiers are encoded and presented at the two designated inputs of the ANN. The forecast results using the 5-bit binary code and then the sine-cosine functions are compared and the MAPE of both methods shown in Table 8.1

Model	Period 1	Period 2
5-bit binary input	1.8630	1.0314
Sine-cosine function	2.2593	1.4296

Table 8.1. Results from the comparison of the encoding methods

From the comparison, the binary method showed better results, suggesting that it is better at encoding categorical day of week information. As a possible explanation for the inadequacy of the sine-cosine method, the day-of-week is not specified as discretely as in the binary method and as such, the days are not classified as accurately.

8.3 Training period

The training period used thus far has been three years. This length of time was deemed long enough to represent most probable trends and short enough to exclude obsolete ones. Once trained, the model was used effectively over a three-month forecast horizon. The major drawback with this method was the time taken to train and verify the model. It was decided to experiment with shortening the length of the training period to determine whether similar or better forecasts could be made, whilst reducing the training time.

An alternative method was considered that involved using a considerably shorter training window length. Simply shortening the window and forecasting over the same time horizon (3 months) was considered inadequate, as the training data set would not contain enough information to predict accurately. Instead, the shorter training window was used to forecasting for a week ahead then moved along one week, retrained and the same process repeated. For the purpose of this discussion, this method is known as the moving training window. A three-month training window was chosen arbitrarily for the comparison with the previous static three-year training period. The assumption behind this method was that all the relevant information for predicting was contained within the most recent weeks and months.

The comparison of the forecast from both models was done over the 12-week period and the MAPE for both methods shown in Table 8.2

Model	Period 1	Period 2
3-year training period	1.8630	1.0314
Moving window	1.0310	0.9909

Table 8.2. MAPE of the comparison between the moving window and static three-year window

The moving window method improved the forecasting accuracy as well as greatly reducing the training time. The improved accuracy is because the models are trained on a cleaner data set without the influence of unwanted noise from past trends. The graphs of the predictions from both models during periods 1 and 25 are shown in figure 8.2a and 8.2b.

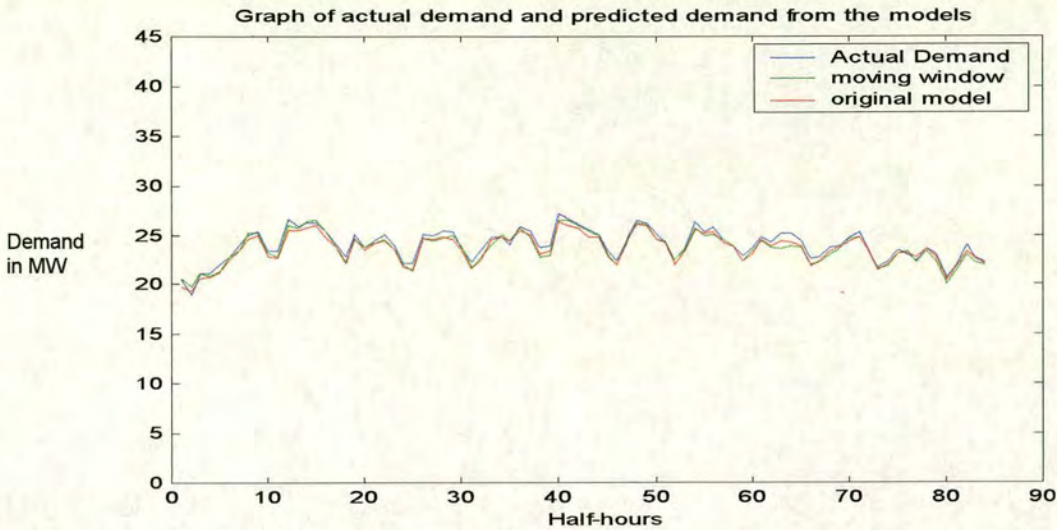


Figure 8.2a. Graphs of the predictions from both models for period 1 over 12-weeks

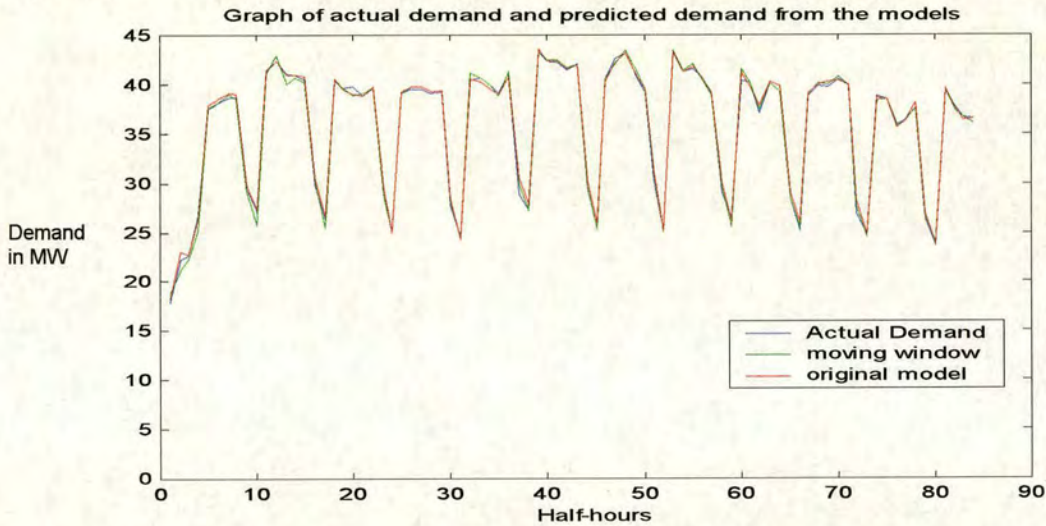


Figure 8.2b. Graphs of the predictions from both models for period 25 over 12-weeks

These graphs show that both models predict demand accurately with the moving window method marginally better for both half-hours. It was thus decided to opt for the moving training window as opposed to the static method.

The next step was to experiment with various window sizes to find the optimal length. However, before doing this, it was decided to modify the inputs to the ANN to make them reflect more realistic NETA balancing mechanism conditions.

8.3.1 Realistic input re-evaluation

From the previous discussions, Suppliers are required to provide the System Operator with final demand notifications at gate closure, which is 4 hours before the scheduled half-hour. This

means that, in practice, demand from the previous 2 hours would not be known at the time of gate closure. Using this method, the model would be required to use forecast data from the previous 2 hours as inputs. If these inputs contain any error, which is highly likely, they would be propagated through subsequent half-hourly forecasts. The complete 48 half-hourly forecasting model would become unstable the further into the future it goes. It was therefore decided to reduce the vulnerability to such errors by decreasing the model's heavy reliance on previous forecast. Instead, the demand forecast from the past hour together with actual demand at gate closure and from the previous week is used. Demand at gate closure was selected because it is the most recent actual demand figure whilst previous week's demand was on the basis of the high day of week correlation discussed in section 6.5.3. For example, the demand at 12:30pm on a given Wednesday will show a strong relationship with demand at 12:30 the previous Wednesday. Though this modification reduces the model's susceptibility to previous forecast error, it is not entirely eliminated.

The predictions over periods 1 and 25 from the revised model were compared with forecasts from the previous example over the first 4 weeks of the 12-week test period and the MAPE from both models presented in Table 8.3.

Previous demand	Period 1	Period 25
Previous 2 hours	1.1504	1.4714
Previous ½, 1, 4 hours and week	1.1770	1.8090

Table 8.3. MAPE of model using previous 2-hour demand and MAPE using previous 1, 4 hour and previous week as lagged demand inputs.

From the Table, the original model outperforms the modified model especially over the marginal period 25. This suggested that during uncertain periods, the ANN requires high quality information nearer to the scheduled half-hour. In spite of this, it was decided to sacrifice accuracy for susceptibility to forecast error. The discussion will be revived in future sections. Contrary to the trends from the previous ANN analysis, the accuracy of both models during period 25 was lower than that for period 1. It was decided to investigate this observation by examining the minimum temperature trend, shown in Figure 8.2 over the 12-week test period. It was noticed that the minimum daily temperature seems to decrease over time. Since the forecast day-ahead minimum temperature determines the teleswitching regime, it might be suggested that over the marginal half-hours, the ANN's predications are better at lower temperatures.

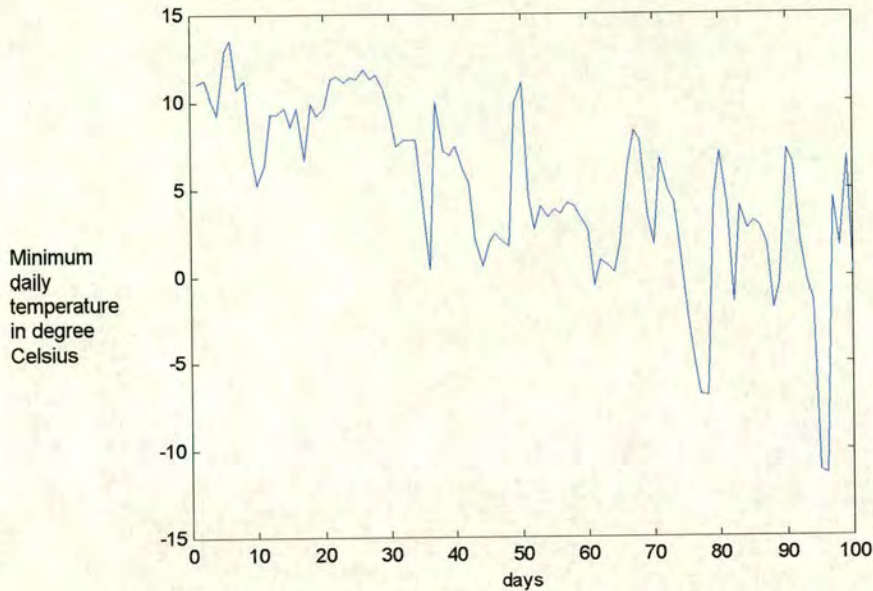


Figure 8.2. Minimum daily temperature over the 12-week period

8.3.2 Optimising the length of the training window

The discussion on training window length digressed in the previous section because it was necessary to define realistic inputs before proceeding with further analysis. In this section various window sizes are experimented with to determine the optimal training length for the neural network. Window sizes ranging from three to six months were compared. Shorter window lengths were disregarded on grounds that they would not contain enough information on a wide range of outcomes. The comparisons were made during periods 1 and 25 over the month used for the previous comparison, the MAPE for the model using various window sizes are shown in Table 8.3.

Window length	Period 1	Period 25
3 month	1.1770	1.8090
4 month	1.0170	1.7122
5 month	1.3790	1.7422
6 month	1.0461	1.6622

Table 8.3. MAPE of the model using different training window lengths

The four-month window length produced the best results for the first half-hour whilst the six-month model produced the best results over period 25. Though the 4-month period required shorter training time, a third half-hour was chosen for a more conclusive comparison. The half-hour chosen was between 18:00-18:30, period 36, because its characteristics are different from all

the others compared in that, it coincides with the evening peak and is also a marginal half-hour. The MAPE are shown in table 8.4

Window length	Period 36
4 month	1.5452
6 month	1.7426

Table 8.4. MAPE of the model using the four and six month windows over period 36

The 4-month window size performs better over period 36 and was chosen as the optimal training period for the complete forecasting model.

8.4 Complete forecasting model

Up to now, this chapter has focussed on optimising the inputs, reducing the reliance on previous forecast and shortening training window size. This section concentrates on the construction of the complete model for all 48 half-hours of the day.

The training and test procedure used up until now was firstly to divide the data into separate training and validation sets, and an independent test set. After training 10 candidate networks with the training and validation data sets, the best is chosen according to its performance on the test data set. This method might not be appropriate in real-life applications, because the model would be required to be trained as close to gate closure as possible. It is therefore preferable not to set aside a separate test set. Additionally, the most effective candidate network should not be chosen according to its performance on the validation data set. This is because generalising performance on the validation data set has a random component due to the noise in the data. The network with the lowest validation error might not produce the best performance on an independent data set.

Various methods exist for choosing the best network, among which is the committee method. Ten candidate networks were trained and validated on the training and validation data sets. All ten models are used to predict demand, the final half-hourly forecast is determined as the average prediction from all the candidate networks. To build the complete forecasts model using this method, 480 such networks had to be trained.

Due to the large number of networks, it was decided to train and test models for the first six hours to determine the validity of this method before proceeding to construct the complete 24 hour version. Figure 8.3 shows the schematic diagram for the first four half-hourly periods of the day it shows that in this method, forecast from the previous two half-hours are used as inputs.

A graph of the forecast is shown in Figure 8.4.

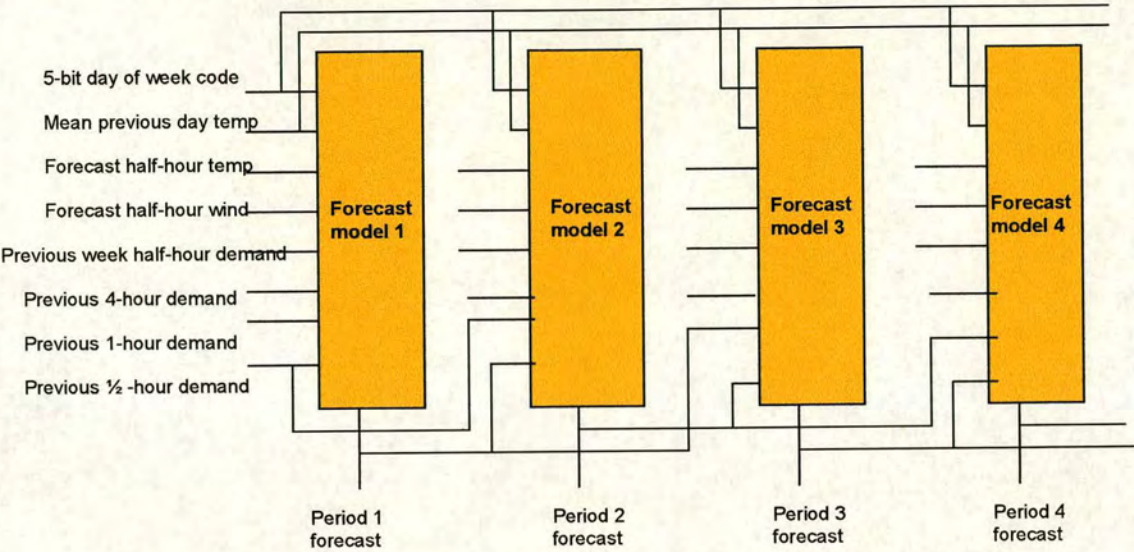


Figure 8.3. Schematic diagram of the first four half-hours of the complete forecasting model

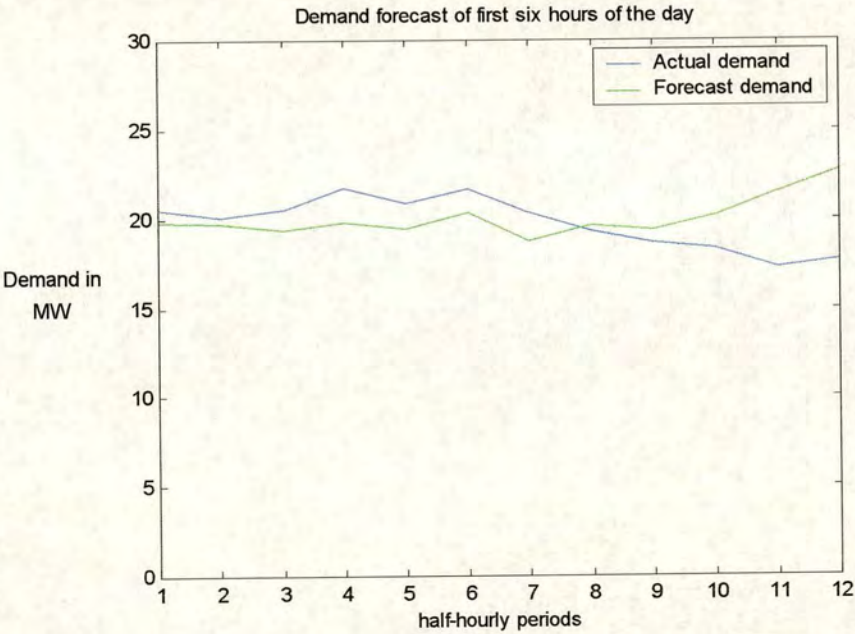


Figure 8.4. Actual and forecast demand for the first 6 hours of the day

The MAPE over the six-hour forecast period is 9.0929. Interestingly, the model follows but under predicts the demand until period 7, after which point the ANN model predicts a steady demand rise, the opposite occurs in reality. From previous discussions, the overnight demand is the most predictable and should thus have the lowest errors. The forecast’s inability to follow the

demand beyond period 7 indicates inherent flaws in the model. A possible explanation is given below.

During training, the back-propagation algorithm assigns weights to the input neurons based on their relative importance to the output. The demand from the previous hour would have the highest correlation with the present demand and would thus be assigned the largest weights. This will mean that any errors in the previous demand forecasts will lead to pronounced errors in the forecast. These errors become multiplicative as they propagate through each half-hourly period. Thus the forecast will diverge further away from the actual demand as it moves into the future. Error compensation methods were considered but discarded because of the complexity in attempting to predict the error. Another important point is that once the model is in operation, there is no process by which it can be recalibrated, so the model will have to be reset. It will not be advisable to use this model on line.

It was decided that though this method would reduce susceptibility to the forecast error, it is still nonetheless vulnerable. Therefore it is inadequate for the purpose of short-term forecasting under NETA due to its susceptibility to errors from previous forecasts.

8.5 Alternative forecasting approach

As the previous approach proved to be inadequate, it was decided to experiment with an alternative method that did not require previous forecasts as input variables. This was done by substituting the previous hour's demand for actual values from the three half-hours prior to gate closure, that is 4, 4 ½ and 5 hours before the scheduled half-hour. Though from the discussion in section 6.5.4, these half-hours will have lower correlation with demand, it was felt that this was a less hazardous trade-off than the uncertainty introduced from previous forecast errors. Though the accuracy of the individual models is sacrificed further, the problem of input errors will be avoided. Figure 8.5 shows a schematic diagram of the inputs to the alternative model.

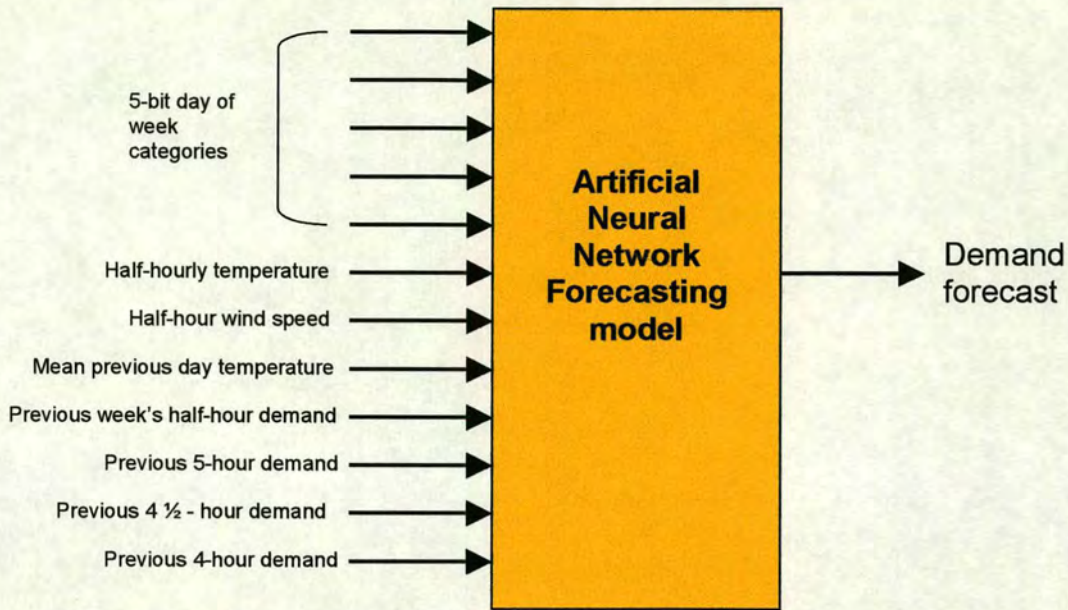


Figure 8.5. Half-hourly forecasting with revised inputs

Since three of the primary explanatory inputs to the model have been revised, it was decided to revisit the analysis on the optimal training window length, to ascertain whether the conclusions drawn in section 8.3.2 still hold. The four and six month window sizes were compared, as they were the optimal from the previous analysis. It was decided not to repeat the comparison of the day of week encoding methods because their information is independent of the other explanatory variables.

The MAPE from both models are presented in table 8.4. Period 12 is included here because it covers the beginning of the morning peak.

Training window size	Period 1	Period 12	Period 25	Period 36
4 month	2.7561	4.8285	3.5264	3.3345
6 month	3.1191	5.2164	3.1908	3.1587

Table 8.4. Comparison of the four and six month window sizes with the alternative model

The 4-month training window was found to be superior in periods 1 and 12 and the 6-month training window superior over periods 25 and 36. The 4-month window was chosen because it was quicker to train. As might be expected, the alternative method shows a noticeable reduction in accuracy.

8.5.1 Network selection: Comparison of the committee and holdout methods

For the six-hour forecasting model in section 8.4, the committee method was used to determine the predictions of each half-hour by averaging the output from ten candidate networks. This method is compared with a variation on the hold-out method discussed in the previous chapter to investigate the relative merits and demerits of both methods for network selection. To select the best candidate network in the hold-out method, the ten models are trained up until one week before the beginning of the intended forecast period. Data from the week preceding the forecast period is used as a test data set from which the best out of the ten candidate networks is selected and used to perform the forecast.

The forecast are performed over a two-week period and the MAPE of the comparison from both models are presented in Table 8.5.

Network selection method	Period 1	Period 12	Period 25	Period 36
Hold-out	3.9449	4.9593	3.4203	2.8645
Committee of networks	3.5267	4.8892	3.3481	2.5639

Table 8.5. Comparison of the committee and holdout method for network selection

The committee method outperformed the hold-out method in all the categories assessed. The higher accuracy with this method can be viewed as arising from the reduced variance due to averaging over many solutions.

After verification of the optimal training length and selection of the best network, it was decided to proceed with the construction of the complete alternative forecasting model. The schematic diagram for the first four half-hours of the complete model is shown in Figure 8.6.

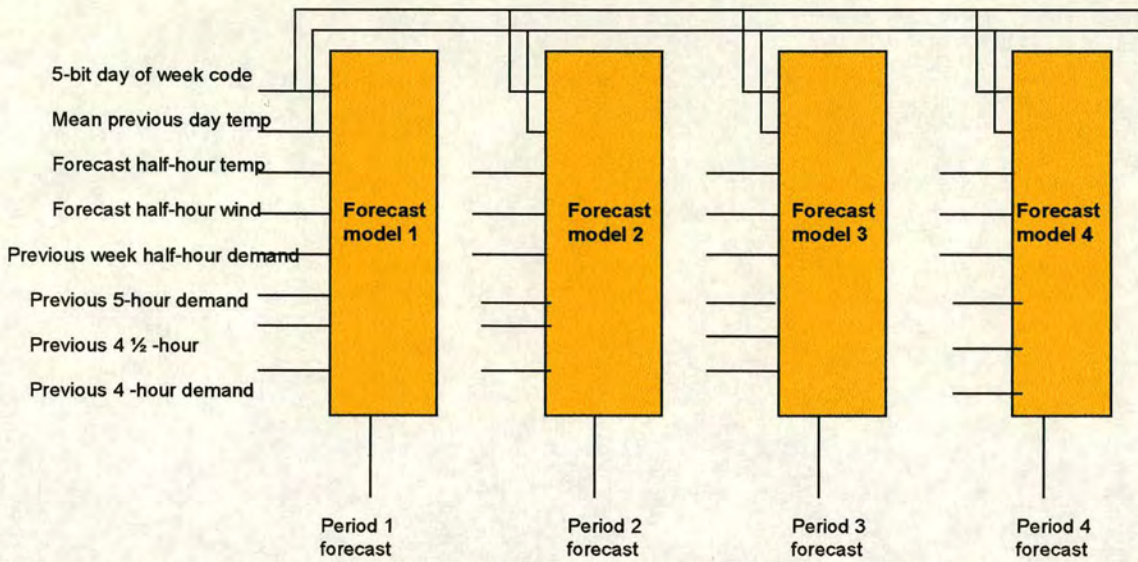


Figure 8.6. First four half-hourly model of the alternative complete forecasting model

48 half-hourly models consisting of 480 candidate networks were built and trained.

8.5.2 Serial (linear) versus parallel forecasting

In section 6.4, where forecasting configuration was discussed, parallel forecasting was chosen as the preferred configuration on the premise that in SHE's Supply area, the time-of-day variations were greater than the day-of-week differences. It was assumed that better predictions could be achieved by assigning dedicated models to learn and predict the demand over each half-hour. To investigate this assumption a complete serial forecasting model was built and the results compared with the parallel forecasting model.

Seven models for each day of the week are built. The inputs to each model are identical to those in the parallel configuration except that the day of week codes are replaced by time of day identifiers. The same binary encoding methodology is used yielding an 8-bit code with three inputs neurons always switched. The schematic diagram for the serial forecasting method is shown in Figure 8.6.

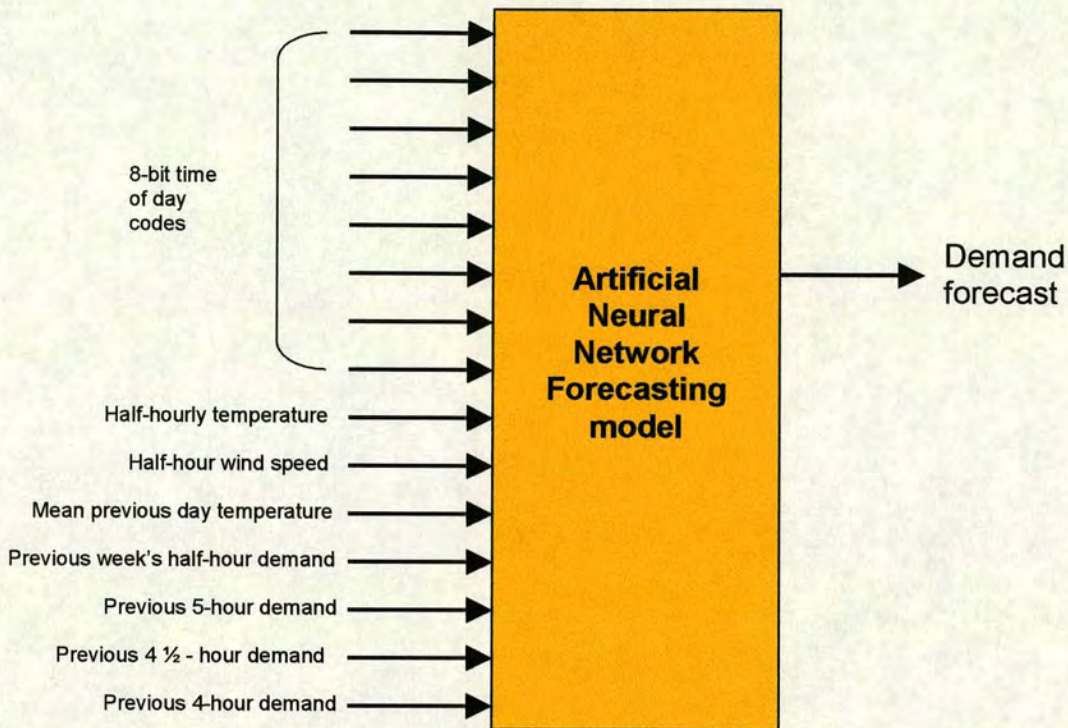


Figure 8.6. Schematic diagram of the serial forecasting for each day

In the linear or serial method the four-month training window is divided among 7 models whereas in the parallel configuration, it is divided among 48 models. A single serial model will thus require more time to train than an equivalent half-hourly parallel model. However, because of the large volume, the complete 48 half-hourly parallel model requires longer for construction and training.

8.6 Results and discussion

Forecasts were performed by both configurations over one week, the predictions from each day are presented and discussed in the following sections.

8.6.1 Monday

Figure 8.7 shows the actual and forecast demand on Monday. From midnight until 18:00 the serial forecast follows the same pattern as the actual demand, albeit raised by a constant magnitude. Beyond 18:00 the forecast is realigned with the actual demand. The parallel method predicts the overnight demand very well. However, between 06:00 and 09:30, the model over-predicts the rate of increase of the morning pick up. This could be because the model assumes that the demand follows the typical weekday morning pattern. Four hours from the start of the

morning pick-up, the model realises that this day is atypical and attempts to correct itself. As such, the predictions are fairly reasonable beyond 11:00.

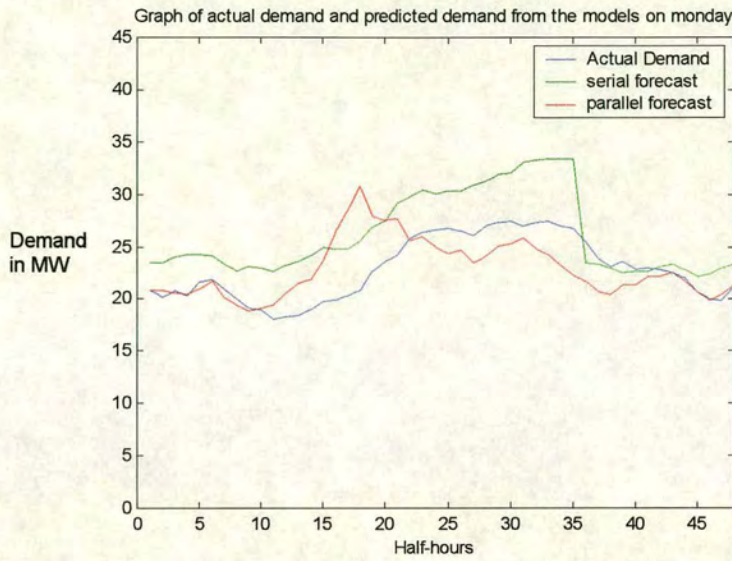


Figure 8.7. Actual and predicted demand for Monday

8.6.2 Tuesday

Figure 8.8 shows the graphs of the predictions from both models for Tuesday. The linear model matches the demand very well on this day. The parallel method produces better forecasts than it did on the previous day, exhibiting comparable performance to the serial method. The improved performance over the morning period could be because Tuesday's demand follows a more typical weekday profile.

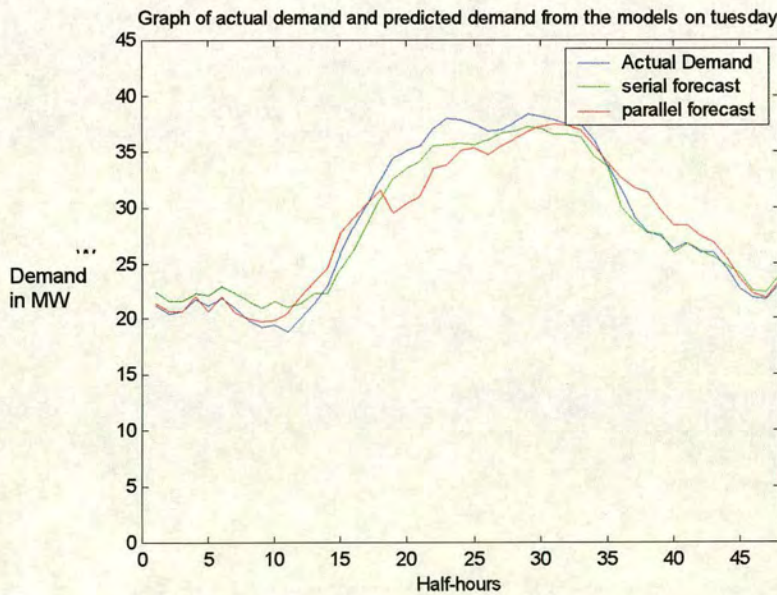


Figure 8.8. Actual and predicted demand for Tuesday

8.6.3 Wednesday

Figure 8.9 shows Wednesday’s demand and the forecast from both models. The serial forecast predicts Wednesday’s demand fairly well however its accuracy is lower than on Tuesday. The parallel forecast follows the demand very closely because Wednesday’s demand shows a very typical profile.

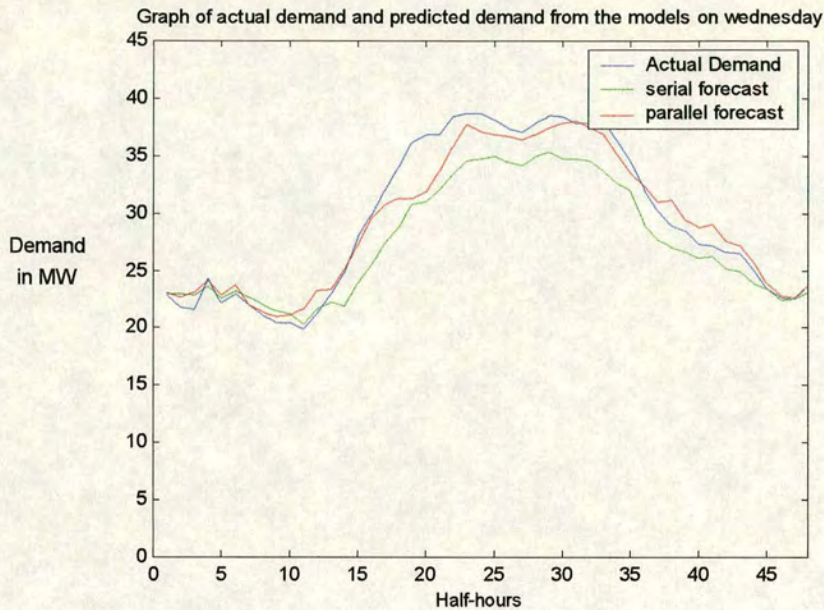


Figure 8.9. Actual and predicted demand for Wednesday

8.6.4 Thursday

Figure 8.10 shows Thursdays demand. The serial model’s performance is worse than on Wednesday. It under predicts the demand with exception of the first five hours of the day. The parallel method forecast’s the demand very well, but it under predicts between 0900 and 1800.

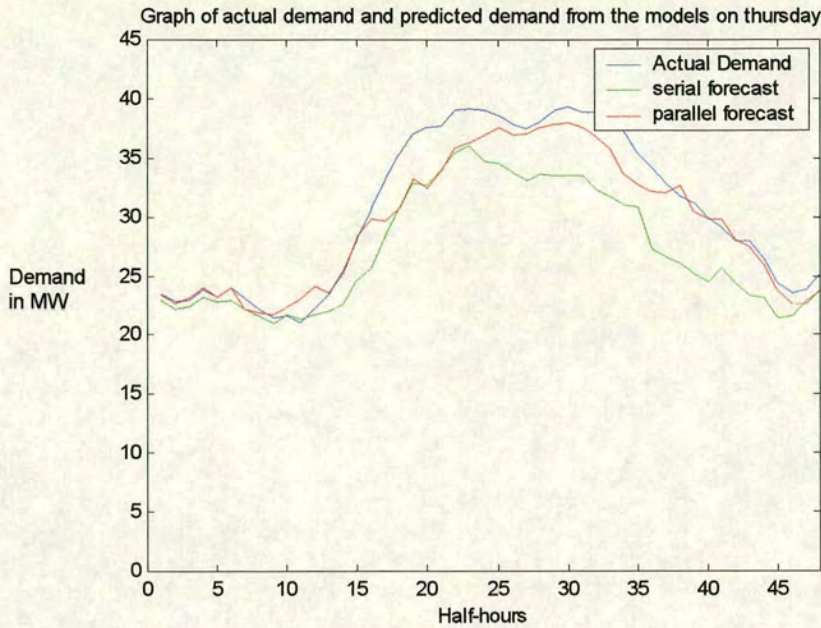


Figure 8.10. Actual and predicted demand for Thursday

8.6.5 Friday

Figure 8.11 shows the demand and predictions on Friday. The demand on Friday is atypical, in that it remains flat for majority of the day with exception of a midmorning demand trough. The performance of both models on this day is of particular interest because the irregular demand pattern examines their generalisation ability to the full.

The serial model predicts the demand fairly well. It replicates the flat demand pattern except between 13:00 and 17:00 where it anticipates a demand peak. Parallel forecasting performs poorly on this day. Similarly to Monday, the model over predicts the demand between 06:00 and 09:30 because it anticipates the morning demand pickup. Once the model is fed with real-information, it attempts to compensate. The forecast becomes aligned to the actual demand between 13:30 and 14:30 and after this point it under predicts. The forecast demand pattern does not resemble the actual profile except after 21:30.

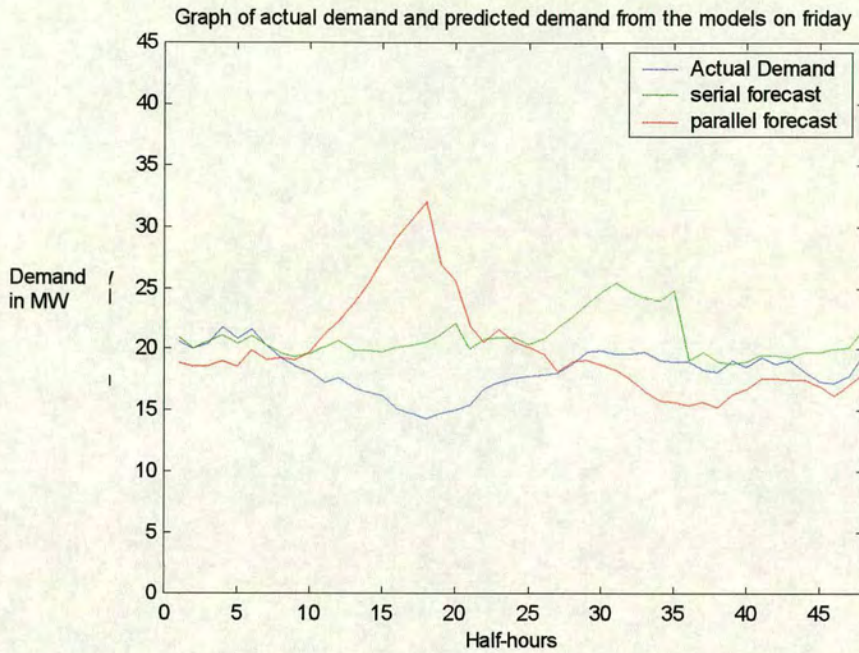


Figure 8.11. Actual and predicted demand for Friday

8.6.6 Saturday

Figure 8.12 shows Saturday's predictions from both models. The actual demand pattern on this day resembles Monday's demand. Though the serial model is less accurate than the parallel model, its demand pattern is closer to the real profile. The parallel model is fairly accurate and gives a good representation of the demand profile.

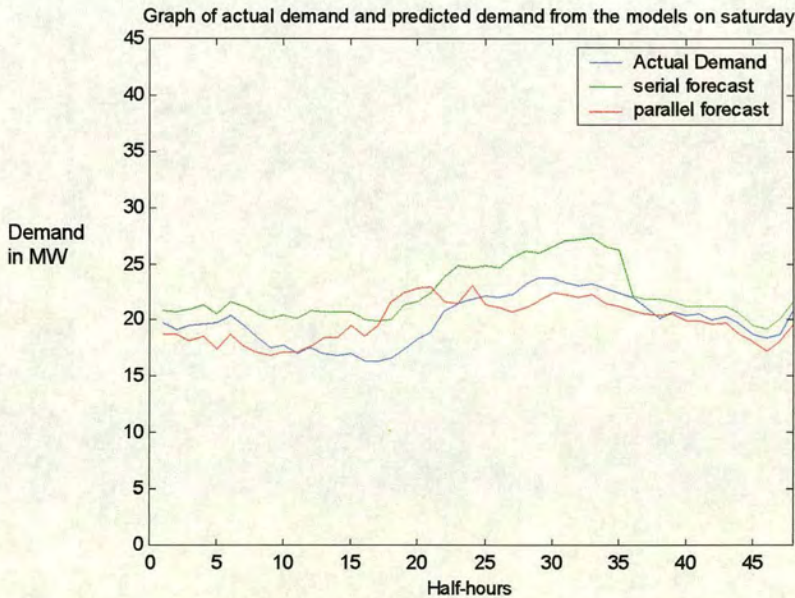


Figure 8.12. Actual and predicted demand for Saturday

8.6.7 Sunday

Figure 8.12 shows the demand on Sunday. Both methods perform very well over Sunday because it is historically very stable, thus predictable.

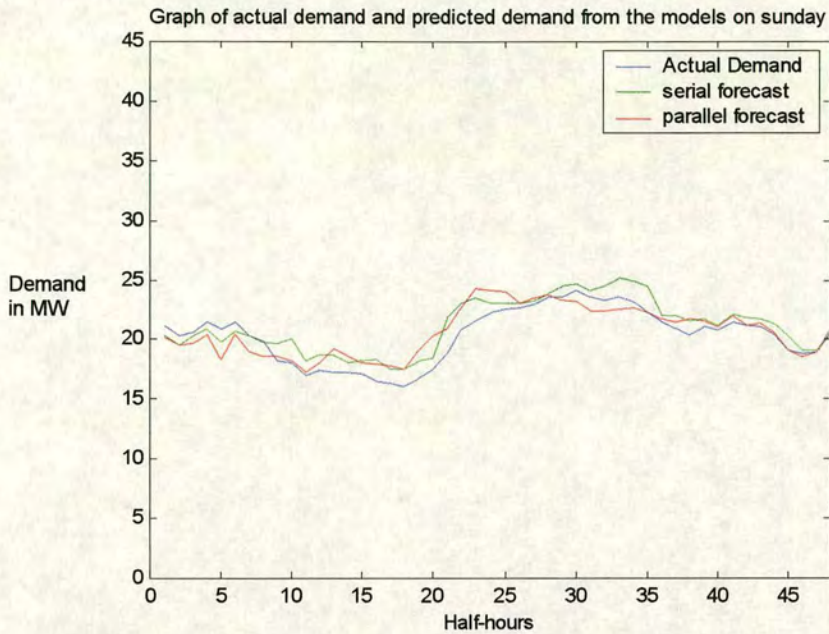


Figure 8.12. Actual and predicted demand for Sunday

8.7 Further discussion

The analysis in the previous section shows that both models are capable of forecasting demand fairly effectively. This section assesses the merits and demerits of both methods by firstly analysing their errors and the predictability of these errors. Predictable errors are less hazardous to Suppliers. Examining the variance of the errors from both models assesses these errors.

8.7.1 Accuracy of models

To examine the accuracy of both models, MAPE is firstly calculated according to the day of the week. This gives an insight into the models ability at forecasting the different daily demand patterns. Secondly, to determine the performance of both models over each half-hour, the MAPE for each of the 48 half-hours is also computed.

8.7.1.1 Accuracy of the day of week predictions

Starting from Friday, the accuracy of both models for each day of the week is presented in Figure 8.13.

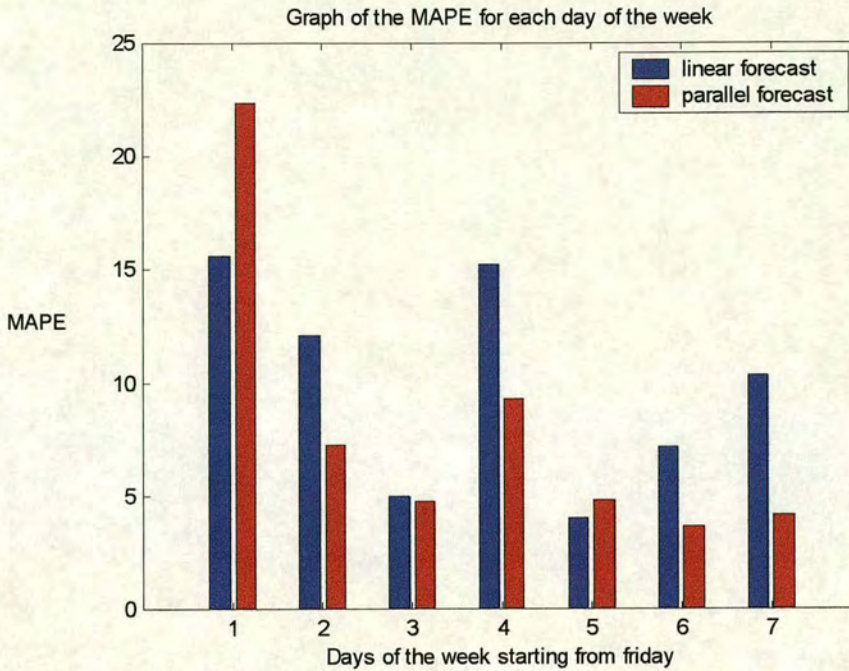


Figure 8.13. MAPE from both models on each day of the week starting from Friday

With exception of Friday and to a lesser extent Tuesday, the parallel method is consistently more accurate. The parallel method's performance on Friday, suggesting poorer generalisation, may imply that the serial model is better at predicting atypical demand patterns.

8.7.1.2 Accuracy of time of day predictions

The MAPE for each half-hour was compiled and presented in Figure 8.14. The parallel model's predictions are generally the more accurate with the notable exception of the morning demand. From the discussion in section 6.5.2.1, the correlation between the temperature, wind-speed and demand is lowest during this period. Thus their explanatory power will be diminished. Additionally, the relevant previous demand inputs are the overnight demand. As is explained in the discussion in section 6.5.3, the overnight demand has little or no relation to the demand during the rest of the day.

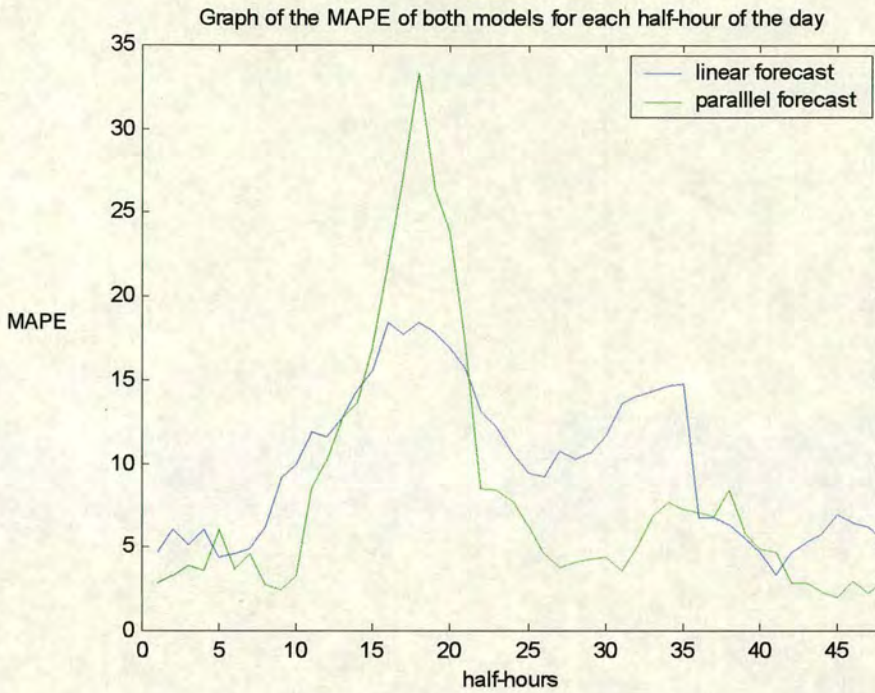


Figure 8.14. MAPE across the 48 half-hours for both models

The combination of these factors means that the models will have difficulty making strong inferences during the mornings. This is especially evident in parallel methods because the models are trained with data from the half-hour they are assigned only. As a result, the network generalises poorly and its behaviour resembles a lookup table. This means that it predicts well when the demand is similar to the typical training set and poorly if there are any deviations. The serial model generalises better over this period possibly because the individual models are trained on a wider, 24-hour data set.

From the previous discussions, each of the seven serial models is trained with a greater number of data points than the 48 parallel equivalents. However, it is not thought that increasing the training window of the parallel forecasting models would improve its generalisation capability because, from the comparison, the four-month outperformed the six-month training window. It could be concluded that window length is not a limitation.

There was no noticeable reduction in MAPE over the teleswitched periods for either model. Thus, it could be assumed that teleswitching is not a major obstacle for either model.

8.7.2 Predictability of the errors

In reality, forecasts are never completely error free, but it is desirable to have predictable errors because they are easily hedged. A simple measure to assess the degree of error predictability is to

compute its variance. Lower variance implies greater predictability and are calculated and assessed for both models in this section.

8.7.2.1 Predictability of the demand shape

Tailoring a hedge to cover the risk of the error is fairly simple if the demand shape is effectively predicted. Both methods ability to predict the demand shape is calculated as the variance of the error for each day of the week over the forecast period, starting from Friday in Figure 8.14. The parallel model's error has higher variance on five out of the seven days with widely differing magnitudes across the week. The serial model's variance was lower on average and stayed fairly constant. These findings imply that the serial model offers better predictions of the demand shape.

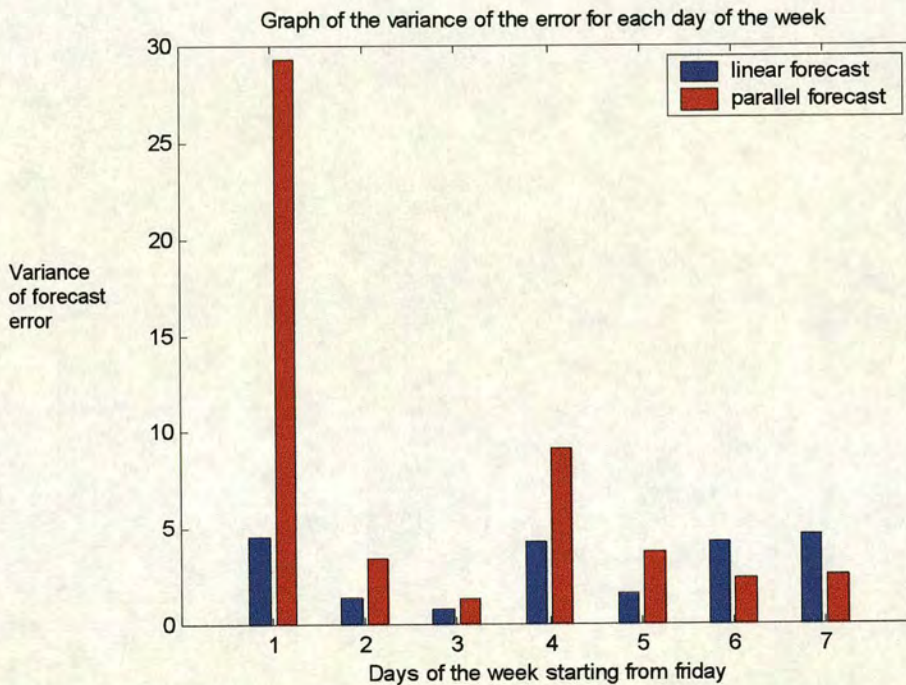


Figure 8.14. Variance of the error for each day of the week starting from Friday

8.7.2.2 Predictability of the error on each half-hour of the day

To assess the predictability of the time of day error, Figure 8.15 shows the variance of the errors for both models across the day. The variance of the error from the parallel method is much lower overall except between 07:00 and 11:00, for reasons discussed in the previous section. On the contrary, the variance of the error from the serial model is more variable across the day. This suggests that with the exception of mornings, the parallel model's time of day errors are much more predictable.

These findings confirm the strengths of both modelling configurations, discussed in section 6.4.

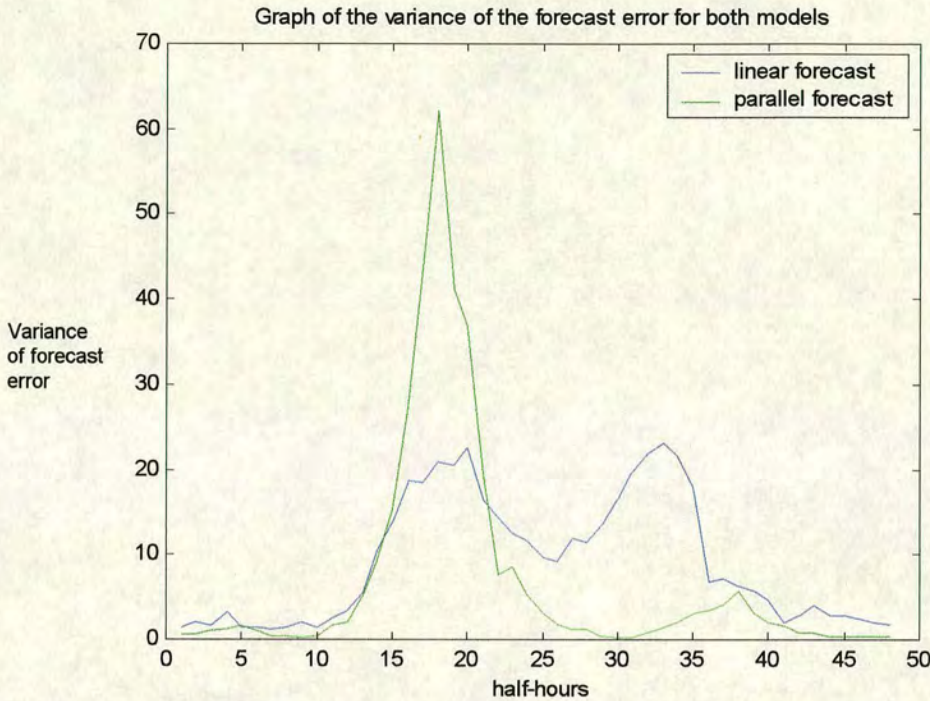


Figure 8.15. Variance of the error for each half-hour over the forecast period

To summarise, the parallel model is more accurate than the linear model, but the serial model is better at predicting the demand shape especially when the demand profile is atypical. It also has the added advantage of being easier to conceptualise and faster to train. Neither model can be said to offer complete advantage over the other.

8.7.3 A simple risk management strategy

From the preceding discussions, both methods have their unique strengths and weaknesses. As a conclusion to this section a simple risk management strategy is presented for each model to demonstrate their application in practice.

8.7.3.1 Risk management strategy using the parallel method

The weaknesses of the parallel method lay in its inability to predict the day of week demand pattern, especially the morning demand. A possible risk management strategy would be to buy Electricity Forwards Agreements (EFA) as discussed in section 4.8.2, to cover the morning demand strip. This will insulate the Supplier during the parallel model's most uncertain period. It is believed that once the morning uncertainty has been hedged against, this model's day of week variations will pose less of a risk to the user.

8.7.3.2 Risk management strategy using the serial method

Though the serial method has higher errors on average, its strength lies in its ability to predict the demand shape. A likely risk management strategy would be to use demand driven swing options as discussed in section 4.8.3.2, in conjunction with the model. Suppliers could inform the option counterparty of the demand shape at gate closure, this would be nominated either upward or downward to compensate for over or under predictions of the demand shape only.

8.8 Case study

In this final section, a case study using real balancing market prices is provided to demonstrate the financial implications of both models. These are contrasted with the fortunes of the naïve forecast described in 7.5, to illustrate the potential savings attainable by using effective predictive tools. The balancing and settlement procedure is very complex. This analysis aims to provide a simple illustration only and is by no means definitive. It was performed using the one-week forecasts from the models.

It is assumed that forecast demand is perfectly hedged as such, for an under prediction, the Supplier is charged System Buy Price (SBP) for shortfalls between the predictions and the actual demand. Conversely, it is assumed that Suppliers can sell the difference between any over prediction and actual demand into the market at System Selling Price (SSP). The models performances are summed up at the end of every day and shown in table 8.6. All the values are expressed in Pounds Sterling. A negative sign represents a payment to the System Operator (SO) and a positive sign represents cash receipt from the SO. The data used to train and test the models was acquired before NETA's inception and as such, the half-hourly SSP and SBP used in this example are taken from an arbitrary week, between Friday 14th September 2001 and Thursday 20th September 2001. These half-hourly prices are shown in Appendix 2.

Model	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	TOTAL
Serial	+1270	+1080	+245	+514	-777	-6178	-4896	-8742
Parallel	-765	-740	-45	-978	-747	-2230	-1002	-6507
Naive	+3593	+469	-453	-3187	-9643	-9598	-3475	-22294

Table 8.6. Imbalance prices from the three models over the one week forecast period

The results show that the users of all three models lose varying amounts of money. The Serial model makes money for the Supplier between Friday and Monday, however its overall losses sum up to £8742 over the week, compared to £6507 from the parallel model. In the broader context, assuming that the losses are consistent over the course of the year, then exposure to imbalance prices will be 52 multiplied by the weekly sums of money at one BSP. Multiplied across all the BSPs controlled by a Supplier is a significant reduction in exposure over a year. For example the

losses a Supplier controlling 15 BSPs will incur using each of the three methods are shown in Table 8.7. This calculation illustrates the financial benefits to Suppliers of effective forecasting methods.

Model	Losses incurred over 1 year
Serial model	6.82 million
Parallel model	5.08 million
Naïve forecast	17.4 million

Table 8.7. Losses over one year from the three models of a Supplier controlling 15 BSPs

So far NETA's participants have been shielded from the total cost of their imbalances. Instead, real imbalance charges have been calculated as an industrial average weighting of the reported SBP and SSP. This means that in practice, participants are charged much less than the quoted prices. At present participants are rewarded for achieving better predictions than the industrial average, this is no mean feat, considering the performance of the naïve forecast.

The volatility of the Balancing Mechanisms prices has been on the increase since NETA's inception and is expected to increase further towards the end of the year, as the margins between supply and demand tighten because of the winter demand. This would transpire into higher more volatile imbalance prices.

It is the author's opinion that either of the forecast configurations discussed would provide substantial financial benefits to their users. The choice of forecast configuration should be based on the trade-offs between accuracy and man-hours used to construct both models.

8.9 Conclusions

This thesis describes the development of a risk management strategy for Suppliers under NETA. As at the time of writing, the markets were still in their embryonic phase so it was decided to focus on the facet of risk that stayed constant regardless of the market structure i.e. volumetric risk. Consequently the author concentrated on building the most effective forecasting methodology to quantify volumetric risk to help insulate Suppliers from imbalance charges.

Neural Networks emerged as the most effective approach for forecasting short-term demand, because they were capable of representing the nonlinearities within the demand. However, after the Neural Network had been optimised and configured to fit a realistic NETA environment, the accuracy fell, though, the reduction was considered not to be a of a sufficient magnitude to merit a rethink of the approach.

Serial and parallel forecasting methods were compared to determine the most effective configuration for forecasting. Neither method was conclusively superior to the other. Though the parallel method was more accurate, it had difficulty predicting morning demand especially on atypical days such as holidays. This difficulty was due to the weak correlation between temperature, wind speed, overnight demand and demand during the morning. It meant that the network functioned as a lookup table during this period. This highlighted a limitation of the approach in that it produced a poor forecast on atypical days. The serial method, on the other hand, required less time to build and train, predicted the demand shape better and also produced better generalisations on atypical days. This was probably because the constituent models were trained over a wider data set. Using this model, Suppliers would find it easier to tailor hedges because the predicted demand shape is closer to that in practice and requires less time to build and operate. Using a case study it was demonstrated that the forecasting approaches would bring substantial financial benefits to their users.

The first part of the thesis of this study was satisfied because the project demonstrated that it was possible to forecast accurately at a BSP supplying a load contaminated by active demand side-management to suit NETA's requirements using an ANN with weather variables among the predictors. The second part of the thesis, which stated that predicting according to the time-of-day should be more desirable than according to the day-of-week, was not proven.

8.10 Summary

This chapter concentrated on completing the analysis and discussion of this thesis by optimising, then building and testing 48 complete forecasting models for NETA. The first half concentrated on improving the efficiency of the ANN method recommended in previous chapter as well as tailoring the models to NETA's requirements. The second part focussed on building 48 complete half-hourly models and performing a case study to illustrate the practical application of the methods developed.

The first section started by comparing the 5-bit encoding methodology employed for categorising the day of the week with an alternative sine-cosine approach. The former method was found to be superior because it did not attach unnecessary weights to any of the days. Next the static three year training period was compared with a dynamic moving training window in an attempt to reduce the training time whilst increasing accuracy. Both goals were achieved and the optimal window length was found to be four months.

Under NETA, gate closure is four hours before the scheduled half-hour, this meant that the models required previous forecasts as their lagged demand inputs. This could introduce errors

with potentially savaging consequences into the model. It was decided to reduce the model's dependency on previous forecasts by substituting the 2 and 2 ½ hour lagged demand inputs with actual values from the demand at gate closure and the previous week. To test the validity of these inputs, a shorter six-hour version was constructed. Forecast error was shown to still be a problem and it was therefore decided to sacrifice accuracy for certainty by substituting the lagged 1 and 1 ½ hour demand inputs with real demand data from the three half-hours prior to gate closure.

In the second part of the chapter, a complete model for the 48 half-hours is constructed. To test the assumption that parallel methods are more effective, the results were compared with those from a complete serial model constructed with the similar inputs. The forecasts were performed over a one-week period. The parallel method proved to be the more accurate of the two however it performed worse during the mornings and whenever demand deviated from its typical value. This suggested that the neural networks linear method made better generalisations. The errors from the linear model were also less variable suggesting better predictions of the daily demand profile. Next, simple risk management strategies were presented to demonstrate the application of either model in reality.

In the final section a case study was performed to illustrate the financial benefits of the methods developed. This was done using balancing market SSP and SBP from an arbitrary week in September 2001. These were contrasted with the fortunes of a naïve forecast. The results revealed that the methods developed reduced the financial risk exposure dramatically, with the parallel method producing superior performance.

Though the parallel method has the advantage of higher accuracy and more predictable time of day errors, it is very labour intensive because it requires 480 candidate networks to be built and trained to forecast the demand every week as opposed to 70 from the serial equivalent. The serial method also offered the advantage of better generalisations and more accurate day of week demand profile predictions, making hedging easier. It is the author's opinion that the choice of model should be based on the trade-off between fewer man-hours and greater hedging flexibility against accuracy, which the Supplier finds acceptable.

9. Conclusion and Summary

9.1 Overview

The challenges facing the industry have changed since the era of privatisation. This has been accompanied by a shift of emphasis from technical to commercial. This project has focussed on the commercial risk within the ESI under the New Electricity Trading Arrangements (NETA). NETA is aimed to encourage a more competitive and efficient trading arena necessitating participants to evolve more sophisticated risk measurement and management techniques to survive. This chapter concludes the discussion on commercial risk management in the ESI by presenting the main arguments in this thesis.

9.2 Implications of NETA

In 1997, it was decided that the electricity trading arrangements centred on the Pool, had not succeeded in delivering lower prices and improving efficiency. The government instructed the Electricity Regulator to come up with new proposals for electricity trading and these were subsequently introduced in March 2001. The aims of these proposals are to increase the efficiency of the industry whilst delivering lower prices to the consumer. To achieve these, the system of trading was changed from a centralised marginal pricing system to a series of bilateral markets known as NETA.

In the NETA's bilateral markets, supply and demand will be matched closer to operating time. This coupled with the abolition of capacity payments could potentially increase price volatility. The introduction of the balancing markets was aimed at rewarding flexible capacity and penalising ineffective participants through punitive imbalance prices.

9.3 Risk and its management under NETA

The nature of ESI risk will evolve under NETA. Credit risk might increase as participants could be expected to experience more volatile income flows. This risk was not a major problem under the previous system because the Pool acted as counterparty to all trades. At present it is difficult to predict the evolution of its management as the markets are still in their infancy and major credit problems have not yet become apparent.

Basis risk is expected to increase, as there will be a wider array of products for participants to choose from, increasing the scope for product mismatches. Time basis should stay the same as it

is at present. Locational basis could become a big problem if balancing mechanism prices evolve to reflect local demand constraints. Operational risk is expected to increase in the early days of NETA as companies adjust their systems to the new trading arrangements. It is also expected that the price volatility in the new markets would result in greater price risk. On the other hand, volumetric risk should stay the same regardless of the market structure, however its financial implications are expected to increase because of the introduction of punitive imbalance prices for out-of-balance market participants.

Previously there was no great need to quantify volumetric risk because its financial implications were negligible. Accurate demand forecasts will be essential to quantify volumetric risk under NETA so as to help avoid imbalance charges. Supplier's previous functions had been primarily billing and metering so they would require a swift evolution in order survive NETA's uncertainties. It was decided to develop an effective short-term forecasting methodology to fit NETA's requirements.

Risk measurement and management are expected to evolve from the instruments available at present. Currently, it is impossible to envisage the long-term nature of price risk because of the limited data. Basis risk will also be difficult to quantify until the true nature of the markets become apparent.

9.4 Forecasting Approaches

The system operator, National Grid Company (NGC) had traditionally performed the forecasting function within the ESI for scheduling, maintenance and long-term system planning. Suppliers forecasting requirements under NETA will be very different as they are required to provide accurate demand positions, Final Physical Notifications (FPN), for every BSP they service, 4 hours before every scheduled half-hour. This will shift the industry's forecasting requirements from long and medium term to short-term. Additionally instead of forecasting centrally at a national level, they will be performed by individual companies for specific locations.

Before accurate demand forecasts are made, the demand has to be understood together with the factors influencing its behaviour. These are referred to as the fundamental demand drivers and were found to be weather variables, time factors, economic, and random factors. From previous studies it was found that the most influential weather variables on demand were temperature and wind-speed. They affected the level of the space and water-heating load, which in some cases contributed to over 75% of the total domestic electricity demand. The time factors included the time-of-day, day-of-week and season. As work patterns influence both domestic and commercial demand, they in turn affect the time of use and the demand on the different days of the week.

Economic and random factors are omitted from short-term forecasting models because of the long time periods over which they occur and then unpredictability.

Once the fundamental drivers of short-term forecasting had been identified, it was decided to investigate the various forecasting methodologies available to select the most appropriate for analytical comparisons. The methods compared were Regressions (parametric and non-parametric), time-series methods, end-user models and Artificial intelligence methods. Regression, notably the Multiple Linear version and Artificial Neural Networks, from the Artificial Intelligence family were chosen. The reason for selecting these are: MLR's are the most widely used forecasting method and their simple linear representation of the demand relationship provides a contrast to ANN, the most popular AI method and also the most effective for modelling a 'cause-effect relationship.

9.5 Data issues

To minimise risk across BSPs Suppliers will have to go with data that exists. The data selection is considered to be very important because the forecast is only as good as the quality of its inputs. The location for the analysis was chosen as Scottish Hydro-Electric's (SHE) 33kV Bulk Supply Point (BSP) at Dyce, in the Aberdeenshire area, servicing a predominantly domestic load. Domestic demand causes the most variability to Suppliers and as such, a domestic load will be more difficult to predict and would provide a more stringent test of the model's capabilities. Additionally, Dyce was chosen because a robust demand data set was easily available and its proximity to the weather station serving the Grampian region suggested that the weather variables obtained would be an accurate measurement of those influencing the demand. SHE had installed active demand-side management, operated by radio teleswitches, to control the heating load. The majority of the heating load was served on an overnight economy tariff and the rest was controlled at various points in the day by teleswitching, the forecast day-ahead temperature determined the duration of these load blocks. The demand side management added additional non-linearity into the relationship between the demand and the weather variables, temperature and wind speed. As a result, this BSP will pose greater challenges to the forecasting models than most in the UK and should expose any limitations in the modelling ability of the chosen models.

Short-term forecasting is far too complex to be performed by a single model. In practice it is either performed by assigning an individual model to learn and predict the demand for each half-hour of the day or each day of the week. For the purpose of this thesis, these are known as parallel and serial or linear forecasting with the former requiring 48 half-hour models and latter, seven day of week models. Parallel forecasting assumes that the intra day variations are greater than the inter day variation whereas linear forecasting assumes the opposite. There had been no definitive studies done to determine the most appropriate method. The majority of examples

identified utilised the parallel approach. For the purpose of the model comparisons, it was decided to use parallel forecasting because the intra day variations seemed greater than the inter day variations. This assumption was to be tested after the preliminary comparisons had been made.

To increase the quality of the inputs, it was decided to use statistical methods to assist in their selection. The weather variables, notably temperature and wind speed were correlated with demand. It was found that the demand exhibited a strong inverse relationship with temperature. Though wind-speed alone did not show a strong relationship, it is acknowledged that wind speed has a marked effect when combined with low temperatures. SHE has quantified this effect within the calculation of the duration of the heating teleswitched blocks. Correlation analysis showed that there was a strong relationship between present demand and demand from the previous two hours. These, combined with the forecast half-hourly temperature, wind-speed and average temperature from the previous day, were selected as inputs to the forecasting model. As it had been decided to forecast according to the time-of-day, it was important to include day-of-week identifiers within the models to categorise the inter day patterns.

9.6 Most effective approach

Using the inputs specified in the previous section, the MLR model was constructed using dummy explanatory variables to categorise the different days of the week. For the ANN, a three layered Multi-layer perceptron was trained using a back propagation algorithm. Both models were trained over a three-year period and used to forecast over a 90-day horizon. Instead of constructing models for each half-hour, periods 1 and 25 were selected because they characterised the load. The results showed the ANN to be superior to the MLR especially over period 25. This was because it was capable of representing the nonlinearities within the demand. The predictions from both models were compared with those from a naïve model (representing the industry standard). Both showed superior improvements suggesting that mathematical forecasting models are much better than the average methods in use at present. Seasonal variables did not seem to make a difference to the forecast accuracy.

9.7 Optimising the ANN

After the comparison, it was decided to optimise the chosen method, the ANN, to achieve greater accuracy whilst shortening the training time. This was done by comparing the model's performance using different approaches for encoding the inter day inputs; using an alternative method for the training window size and; comparing two approaches for selecting the best candidate network. It was found that a binary method using the lowest bit code with the same number of bits switched for each day of the week was the most effective encoding method. The

shorter training window size (with forecast performed over a week, then the window moved along one week) was found to be the most effective because it reduced training time and produced more accurate results, as the model was trained on a purer data set. The optimal training window length was found to be 4 months. For candidate network selection, the more effective approach was found to be the committee method. Ten candidates were trained for each half-hour and the forecast was determined as the average of their predictions.

One of the principal predictor variables used for the comparisons was lagged 2-hour demand. In practice, demand expectations will be submitted to the System Operator at Gate Closure, 4 hours before the scheduled half-hour. This will mean that the lagged demand will be forecasts because the actual values are not known four hours in advance. Any errors, which are highly likely, would be propagated through the subsequent forecast. This is highly undesirable, as it would make the model very unstable. It was decided to reduce this vulnerability by using 1 hour lagged demand, the demand at gate closure and the previous week's demand. In spite of this, the model still showed instability. It was decided that the stability was more desirable than the reduced accuracy and as a result, the 1-hour lagged demand variables were replaced with actual demand values from the three half-hours prior to gate closure as well as the demand from the previous week.

9.8 Parallel versus serial (linear) forecasting

A forty-eight half-hour parallel model had been constructed as it was assumed that this was the most effective configuration for forecasting demand at the BSP. Comparing the forecasts with those from a serial forecast tested this assumption. This was done over a week in March 1999. Neither method was conclusively superior. The parallel method was more accurate but it had difficulty predicting morning demand especially on atypical days. The difficulty in generalising was due to the weak correlation between the temperatures, wind-speed, overnight demand and demand over this period. This meant that during this period, the network functioned as a lookup table, highlighting a limitation of the approach. The serial method, on the other hand, required less time to build and train, predicted the demand shape better and also produced better generalisations on atypical days. This was probably because the constituent models were trained over a wider data set. It is the author's opinion that the choice of configuration should be based on the trade-off the Supplier finds acceptable.

9.9 Application of approach

Both the linear and parallel models have their relative strengths. Either could be applied in conjunction with a risk management strategy to measure and manage the exposure to imbalance prices. An example of risk management strategies using either model is presented below.

- Parallel model: The Supplier could buy Electricity Forward Agreements to cover the morning demand strip.
- Serial model: It could be used in conjunction with demand driven swing options that could be adjusted up or down to compensate for over or under predictions.

It was decided to compare both approaches with the naïve forecast in a case study with real balancing market prices to demonstrate the financial advantage they would bring their users. Paying Suppliers System Selling Price (SSP) for any over predictions and charging System Buy Price (SBP) for demand shortfalls performed the calculation. The parallel model recorded the smallest losses, £6507, followed by the serial model £8742, with the naïve forecast losing £22294. These findings confirm the superiority of the approaches developed to the industry standard. This provided confirmation for the achievement of the first project aim; to develop effective forecasting methods for NETA.

9.10 Suggestions for future work

This study was concerned with the development of commercial risk management in the Electricity Supply Industry. It focussed primarily on the development of an accurate demand forecasting methodology to help avoid imbalance charges. The work described in this thesis, although complete in itself, has a number of possible avenues for expansion that the limited period of study under NETA did not allow. The scope for future work is extensive as the market is still in its embryonic stage, further applications of the model could become apparent over time.

9.10.1 Managing the volumetric risk

The forecasting model measures volumetric risk. Combining the forecasting methodologies with risk management tools could be used to assess the most appropriate risk management strategy to use in conjunction with each of the forecasting methods.

9.10.2 Predicting and managing the risk of forecast error

From the discussion, it is apparent that no forecast is completely free from errors. A suggestion for future work would be to investigate the forecast errors further by attempting to predict them. This could be done by forecasting from both models over a much longer time period, calculating the errors and using a separate ANN to learn and forecast the error. If this method is successful, it could increase the accuracy of the forecast even further.

Following on from this, would be to manage the risk of forecast error using external hedges. There are a variety of hedging tools available in the market place. A hedged position could be

stimulated using several examples of hedges and their performance compared over a period of time to determine the most appropriate external hedge for managing the risk of forecast error.

9.10.3 Application of models to Generator forecasting

A third suggestion would be to apply the model to forecasting generator demand. Generators also need better forecasting procedures to optimise their dispatch schedule.

9.10.4 Hybrid forecasting

A fourth suggestion would be to use a genetic algorithm to select optimal inputs and outputs to train the ANN. The ANN might be able to make stronger inferences from more optimal inputs and outputs.

9.11 Final conclusions

The volatility of the Balancing Mechanisms prices increased continually since NETA's inception and is expected to increase further towards the end of the year, as the margins between supply and demand tighten because of the winter demand. This could lead to higher and more volatile imbalance prices. It is the author's opinion that either of the forecast configurations discussed could provide substantial financial benefits to their users. The choice of forecast configuration should be based on the trade-offs between accuracy and man-hours used to construct both models. The choice of either forecasting method would help to insulate Suppliers from prohibitive imbalance prices. Though this discussion focused on Suppliers, Generators could utilise the models developed to improve the efficiency of their scheduling.

This thesis confirmed that ANNs are the most accurate method for predicting short-term electricity demand. It achieved its first aim of designing a neural network methodology to forecast demand at a bulk supply point contaminated by demand-side management to suit NETA's requirements. The results from forecasting according to the time of day were not conclusive enough to prove the thesis's second aim to demonstrate that time of day were superior to day of week predictions.

References

-
- ¹ Doyle, G., Maclaine. D., "Power as a commodity: The future of the UK electricity supply industry", Financial Times Energy publishing, 1996
- ² Waddams Price, C. " Competition and Regulation in the UK Gas Industry", Oxford Review of Economic Policy, vol.13, no.1, pp.47-63, 1997
- ³ Whittington, H. W. W., "Fourth year power and machine notes", The University of Edinburgh, 1990.
- ⁴ Silverton, C. L., " Modelling Long-term primary energy mix in the Electricity Supply Industry through Genetic Algorithm Based Optimisation" PhD thesis, The faculty of Science and Engineering, The University of Edinburgh, 2000.
- ⁵ "Privatising Electricity", The Government's proposal for the privatisation of the electricity supply industry in England and Wales, HMSO, 1988
- ⁶ Hoare, J., " The UK electricity market", Managing Energy Price Risk, Chapter 11, Risk Publications, London, 1995.
- ⁷ Green, R.J., Newbury, D.M., " Competition in the British Electricity Spot Market", Journal of Political Economy, Vol. 100, no.5, pp.929-53, 1992.
- ⁸ " Review of Scottish Trading Arrangements: A consultation document ", The Office of Gas and Electricity Markets, Birmingham, October 1999
- ⁹ " The Energy Report 1999", The Department of Trade and Industry, HMSO, 1999
- ¹⁰ "Ofgem Annual Report: Protecting Customers", The Office of Gas and Electricity Markets, Birmingham, May 2000.
- ¹¹ OFFER, "Report on pool Price Increases in Winter 1997/98", Birmingham, Office of Electricity Regulation, June 1998
- ¹² Newbury, D.M., " Power market and Market Power", The Energy Journal, Vol.19, no.3, pp41-66, 1995.
- ¹³ Green, R.J., Newbury, D.M., " The Electricity Contract Market in England and Wales" Journal of Industrial Economics, forthcoming, 1998
- ¹⁴ Keefe David, " Pool Wars", Energy Risk, Vol2, no7, July 1995
- ¹⁵ Wooley, J., " Power Games" Energy Risk, vol.1, no. 4, May 1994
- ¹⁶ " Review of Electricity Trading Arrangements: proposals", Office of Electricity Regulation, Birmingham, July 1998.

-
- ¹⁷ “The New Electricity Trading Arrangements”, The Office of Gas and Electricity Regulation, Birmingham, July 1999.
- ¹⁸ Halldearn, D., “The Reasons behind the New Electricity Trading Reform”, Electricity Trading, IEE Power Seminar, London, 6th March 2000.
- ¹⁹ “The New Electricity Trading Arrangements: A draft specification for the balancing mechanism and imbalance settlement”, version:1.2, The Office of Gas and Electricity Markets, July 1999
- ²⁰ “Interim Proposals for the reform of Scottish Trading Arrangements: British Electricity Trading and Transmission Arrangements (BETTA)”, The Office of Electricity and Gas Regulation, Birmingham, August 2000
- ²¹ “Review of Scottish Trading Arrangements: A consultation document”, The Office of Gas and Electricity Markets, Birmingham, October 1999
- ²² Wolfram, C. D., “Electricity Markets: Should the Rest of the World Adopt the UK Reforms?”, University of California Energy Institute, POWER working papers, PWP-069, September 1999
- ²³ Green, R., “Draining the Pool: The reform of Electricity Trading in England and Wales”, Energy Policy, forthcoming, December 1998.
- ²⁴ “Role playing simulations of the new electricity trading arrangements”, The Office of Gas and Electricity Markets and London Economics, October 1999
- ²⁵ Hoare, J., “Financial Derivatives and Risk Management”, GNI Ltd, Issue 9, The UK electricity market, April 1997
- ²⁶ Waters, L., “Implications and Opportunities for Customers”, Electricity Trading, IEE Power Seminar, London, 6th March 2000.
- ²⁷ ‘DEFRA: Climate Change UK Programme’, www.defra.gov.uk/environment/energy/index.htm
- ²⁸ Doyle, G., Maclaine, D., “Power as a commodity: The future of the UK electricity supply industry”, Financial Times Energy Publishing, 1996.
- ²⁹ ‘Managing Business Risks’, Economist Intelligence Unit, London, 1995.
- ³⁰ Cummins D., Philips D., Smith S.D., ‘The rise of risk management’, Federal Reserve Bank of Atlanta Economic Review, First Quarter 1998.
- ³¹ Charette R.N., ‘The Risks with Risk Analysis’, Communications of the ACM, Vol. 34, Issue 6, 1991, pp106
- ³² Pollatsek A., Tversky A., ‘A Theory of Risk’, Journal of Mathematical Psychology, Issue 7, pp 540-553
- ³³ Holton, G.A., ‘Closed form value at risk’, www.contingencyanalysis.com
- ³⁴ Dybvig P.H., Marshall W. J., ‘The New Risk Management: The good, the bad and the ugly’ Federal Reserve Bank of St Louis Review, Nov/Dec 1997
- ³⁵ ‘Credit Risk : special report’, Energy and Power Risk Management, Vol. 5, No 2, May 2000

-
- ³⁶ Hampton M., 'Options', Managing Energy and Price Risk, Risk publications, London 1995, pp 97-115
- ³⁷ Martin J., 'The pool price is a furphy' Energy and Power Risk Management, Vol. 3, No 6, October 1998
- ³⁸ Prem Chandran, Personal communications, 2nd Annual Industry Forum on Accurate Electricity Demand Forecasting, Stakis St.Ermin's Hotel, London, 16th and 17th May 2000
- ³⁹ 'Weekly Pool exposure Report', Scottish-Hydro, Perth, 4th December 1998
- ⁴⁰ Jenny Lees, Personal communications, Scottish-Hydro, Perth, June 1998
- ⁴¹ Spector Y., Tishler A., Ye Yinyu, 'Minimal Adjustment costs and the optimal choice of inputs under time-of-use electricity rates', Management Science, Vol. 41, No. 10, October 1995
- ⁴² 'Load Research progress Report', www.electricity.org.uk.
- ⁴³ Allera S.V., Horsburgh A.G., 'Load profiling for energy trading and settlements in the UK electricity markets', Load Research Group, Electricity association, UK.
- ⁴⁴ Liz Wright, "The history of the CFD and EFA market: some critical comments", Managing Energy Price Risk, pp217
- ⁴⁵ Pilipovic D., 'Energy Risk: valuing and managing energy derivatives', McGraw-Hill, U.S.A, 1998
- ⁴⁶ 'Special report-weather risk', Energy power and Risk Management, Vo'. 5, No 5, August 2000
- ⁴⁷ 'Europe's first weather derivative deal announced', Energy and Power Risk Management, Vol. 3, No 8, Dec 1998/Jan 1999, p 3
- ⁴⁸ Redford, S. J., 'The Rationale for Demand-side Management', IEE Power Engineering Journal, October 1994, pp 211-217
- ⁴⁹ Gellings P. E., 'Demand forecasting in the electricity utility industry', PennWell, Tulsa, Oklahoma, U.S.A, 1996, pp9-2:9-3.
- ⁵⁰ 'How does CHP work?' <http://www.chpa.co.uk/aboutchp.htm>
- ⁵¹ St Clair, Richard, 'Load management technology-future challenges', IEE Electricity Trading Seminar, 00/038, London, March 2000
- ⁵² Elton J.E., Gruber M. J., 'Modern portfolio theory and investment analysis', 5th edition, John Wiley & Sons, U.S.A, 1995.
- ⁵³ Dowd, Kevin, 'Beyond value at risk', John Wiley & Sons, Chichester, England, 1998
- ⁵⁴ 'RiskMetrics-Technical document' www.jpmorgan.com/RiskManagement/RiskMetrics/riskMetrics.html.
- ⁵⁵ Henney Alex, Keers Greg, 'Managing total corporate electricity market risks' EPRI conference, June 1998, Washington DC, U.S.A
- ⁵⁶ Keers Greg, 'Taking the corporate risk out of power trading', www.kwi.co.uk/articles.htm
- ⁵⁷ Dorey H., 'Advanced Metering in old and new worlds', IEE Power Engineering Journal, August 1996, pp 147-148

- ⁵⁸ Jenny Lees, Personal communications, Scottish-Hydro, Perth, August 1998
- ⁵⁹ Gellings C. W., 'Demand forecasting in the electric utility industry', 2nd Edit, Pennwell Publishing, Oklahoma, USA, 1996
- ⁶⁰ 'Digest of UK energy statistics 2000', Department of Trade and Industry HMSO, The government statistical service, 2000
- ⁶¹ 'The Energy Report 1999', The Department of Trade and Industry HMSO, 1999, pp 48
- ⁶² 'An analysis of electricity usage: Domestic lighting', www.electricity.org.uk/services/lrg/light1.html
- ⁶³ Connolly John, Personal Communications, Scottish Hydro Electric, Pitlochry, August 1999
- ⁶⁴ Davies, M. 'The relationship between weather and electricity demand', IEE monograph No. 314S, 1958
- ⁶⁵ Peirson, J.D., Henley, A., 'An analysis of the electricity association's methodology for standardising demand data for temperature', Report for load research group of the electricity association services ltd. Kent Energy economics Research group, 1992.
- ⁶⁶ Henley, A. Peirson, J.D., 'Non-Linearities in electricity demand and temperature: Parametric vs Non-Parametric methods', University of Kent Economics Discussion paper, No. 95/8. 1995
- ⁶⁷ Kreith, F. 'Principles of Heat transfer', 4th Edition., Harper-Collins, New-York, 1986
- ⁶⁸ Peirson, J., Henley, A., 'Electricity Load and Temperature: Issues in dynamic specification', Energy Economics, Vol. 16, No. 4, 1994
- ⁶⁹ Anderson, B.R., Chapman, P.F., Cutland N.G., Dickson, C.M., Shorrock, L.D., 'Bredem-12: Model description', Building Research establishment Laboratory Report, BRE publications, Watford, 1996
- ⁷⁰ Peirson, J., Henley, A., 'Non-Linearities in Electricity Demand and Temperature: Parametric versus Non-Parametric methods', Blackwell Publishers Ltd, Oxford, UK, 1997
- ⁷¹ Artaiz C.J, Soto C.D., Tome M., 'SIPREDE: A medium and long-term electric load forecasting system', Proceedings of the 3rd European Intelligent Forecasting Systems workshop, Santorini, Greece, 7th – 8th June 2000. pp75-78
- ⁷² Ryan, T. P., 'Modern Regression Methods', John Wiley and Sons, Inc. USA, 1997, pp245-251
- ⁷³ Charytoniuk, W., Chen, M.S., 'Nonparametric Regression Based short-term load forecasting', IEEE transactions on Power Systems, Vol. 13, No. 3, August 1998, pp725-730.
- ⁷⁴ Eubank, R. L., 'Spline smoothing and Nonparametric regression', Dekker, New York, 1988
- ⁷⁵ Box, G.E.P., Jenkins, G. M., 'Time series analysis: forecasting and control', 3rd Edition, Prentice-Hall, Englewood Cliffs, New-York, 1994.
- ⁷⁶ Hagan M. T., Behr S.M., 'The time series approach to short term load forecasting', IEEE Transactions on Power Systems, Vol. 2, No. 3, August 1987, pp785-791
- ⁷⁷ Al-Alawi, S.M., Islam S.M., 'Principles of electricity demand forecasting: part 1; methodologies', Power Engineering Journal, June 1996.

-
- ⁷⁸ Small J.L., 'Joint estimation of household and heat-pump electricity demand'', IEEE Transactions on Power Systems, Vol. 9, No. 1, February 1994
- ⁷⁹ Bimal, K. Bose., 'Expert system, Fuzzy Logic and Neural Network applications in Power electronics and motion control', Proceedings of the IEEE, Vol. 82, No. 8, August 1994
- ⁸⁰ Rahman, S., Bhatnagar, R., 'An expert system based algorithm for short-term load forecast', IEEE Transactions on Power Systems, Vol. 3, No. 2, May 1988
- ⁸¹ Rahman, S., Drezga, I., Rajagopalan, J., 'Knowledge enhanced connectionist models for short-term load forecasting', ANNPS '93, proceedings of the second international forum of the application of neural networks to power systems, Yokohama, Japan, April 1993, pp401-407
- ⁸² Zadeh, L.A., 'Outline of a new approach to the analysis of complex systems and decision processes', IEEE systems, Man and Cybernetics, Vol. SMC-3, No. 1, 1973, pp. 28-40
- ⁸³ Ross, T. J., 'Fuzzy logic with engineering applications', McGraw-Hill, Inc, USA, 1995
- ⁸⁴ Liu, K., Subbarayan, S., Shoults, R. R., Manry, M.T., Kwan, C., Lewis, F.L., Naccarino, J., 'Comparison of very short-term load forecasting techniques', IEEE Transactions on Power Systems, Vol. 11, No. 2, May 1996, pp877-882.
- ⁸⁵ McCulloch, W. S., Pitts, W., 'A logical calculus of the ideas immanent in nervous activity', Bulletin of mathematical biophysics, Vol. 5, 1943, pp115-133
- ⁸⁶ Madan, S., Bollinger, K. E., 'Applications of artificial intelligence in power systems'. Electric Power Systems Research 41, Elsevier Science, 1997, pp117-131
- ⁸⁷ Daneshdoost, M., Lotfalian, M., Bumroongit, G., Ngoy, J. P., 'Neural network with fuzzy set-based classification for short-term load forecasting', IEEE transactions on Power Systems, Vol. 13, No. 4, Nov 1998, pp1386-1391
- ⁸⁸ Heng, E.T.H., Strinivasan, D., Liew, A.C., 'Short-term load forecasting using genetic algorithm and neural networks', Energy Management and Power Delivery, Proceedings of EMPD '98, Vol. 2, 1998, pp 576-581
- ⁸⁹ Dick, Alan., 'Effect of competition in supply on demand control in the UK', Electricity Association, UK, November 1997.
- ⁹⁰ Masters, T., 'Practical Neural Network Recipes in C++', Academic Press, San Diego, USA. 1993.
- ⁹¹ Bishop, Christopher M., 'Neural Network for Pattern Recognition', Oxford University Press, Oxford, England, 1995.
- ⁹² Reed, Russell D., Marks II, Robert J., 'Neural Smithing: Supervised learning in Feedforward artificial Neural Networks', MIT press, London, England, 1999
- ⁹³ Fidalgo, J.N., Pecas Lopes, J.A., 'Load forecasting dealing with medium voltage network configuration', Proceedings from the 8th European Symposium on Artificial Neural Networks, Bruges April 26-28, 2000. pp407-412

Appendix 1

Published papers

1. D.Adjepon-Yamoah, H.W. Whittington, A.P. Birch., ‘ Assessing the effect of weather variables on customer demand’, 34th Universities Power Engineering Conference, Vol. 2, p 505-508, Sept 1999.
2. D.Adjepon-Yamoah, H.W. Whittington, A.P. Birch., ‘ Determining the weather uncertainties in domestic electricity consumption’, Proceedings of the 3rd European Intelligent Forecasting Systems Workshop, p85-91, June 2000.

ASSESSING THE EFFECT OF WEATHER VARIABLES ON CUSTOMER DEMAND

D. Adjepon-Yamoah*, H.W. Whittington*, A.Birch**

Department of Electrical Engineering, University of Edinburgh*. Scottish and Southern Energy Plc**

ABSTRACT

The introduction of the new wholesale trading arrangements in autumn 2000 will pose a different set of challenges to the UK electricity industry. With no uniform reference price, as is the case at present, and the lack of historical data, market participants face greater uncertainty over the electricity price volatility. A thorough understanding of the effects of the market fundamentals on demand and generation is needed to give an indication of electricity fair price.

This paper considers the impact of the weather variables on customer demand by comparing the different available methods for determining this relationship. The models are currently being developed.

1. INTRODUCTION

The complete deregulation of the electricity supply industry last year, and the impending, proposed trading mechanism to be introduced in autumn 2000, will result in increased uncertainty for participants in the England and Wales Electricity Supply Industry (ESI).

With the present market mechanism of marginal pricing, all generators receive Pool Purchase Price (PPP) and Suppliers pay Pool Selling Price (PSP). The PPP consists of the System Marginal Price (SMP) and a capacity element, which is the opportunity cost of insufficient capacity. The PSP is made up of PPP and a component for transmission losses and sundry expenses. Under this mechanism, the cost of, and incentive to make excess capacity available are rewarded through capacity payments.

1.1. New trading arrangements

In contrast, the new proposed trading arrangements will be based on an auction market supported by longer-term bilateral contracts. This has several advantages over marginal pricing, including Demand Side Participation (DSP) which is more economically efficient, but introduces higher price volatility.

One main reason for the increased volatility is that, there is no provision for making extra capacity available. In some cases there is an incentive for Generators to tighten margins, so as to increase the cost of Contracts-for-Difference (CfD)(1). The Australia Pools have a very similar structure to the new proposals. The Victoria pool price averaged \$22 per MWh, but

at the margins the Value of Lost Load (VOLL) was about \$22,000 per MWh. These are known as 'needle peaks' (2).

Because the impending changes in the England and Wales Pool will cause much higher uncertainty, resulting in greater risk for market participants, players will need to have an accurate understanding of their consumer consumption patterns to stay competitive.

1.2. Domestic demand

This paper concentrates on the domestic segment of the market, because it is the most variable demand segment. For example, on a given day, no two households will have identical half hourly demand and even within a household consumption patterns vary from day-to-day.

1.3 Fundamental drivers

The fundamental drivers of electricity demand are

- Economic trends
- Timefactor (e.g. weekly, seasonal, holidays)
- Weather factors
- Random effects

Experience shows that weather effects on residential demand are the source of the greatest variability. Certainly, needle peaks in demand, caused by weather effects, have been experienced in the past; usually associated with a cold snap. For this reason the relationship between the weather and electricity demand has been selected as the basis of this study.

It considers four approaches for defining this relationship.

2. EFFECT OF THE WEATHER ON RESIDENTIAL DEMAND

This study identifies as the major uses of electricity in domestic households, the following:

- 1) space heating
- 2) water heating
- 3) cooking
- 4) dishwashing
- 5) refrigeration/freezer
- 6) television/hi-fi
- 7) others(e.g. kettles, vacuum cleaners)

2.1 Weather variables affecting demand

The weather effects found to affect demand are as follows

- 1) temperature
- 2) wind speed
- 3) illumination effects
- 4) precipitation.

The temperature and wind speed affect the level of heating demand: Heat demand depends on the following;

- 1) desired level of thermal comfort
- 2) rate of heat loss from dwelling.

The rate of heat loss from a building in turn depends on the following:

- (i) Quality of insulation
- (ii) Wind speed
- (iii) Duration of the cold/hot spell
- (iv) Rate of heat build-up (dependent on the density of buildings)

Temperature and wind speeds also have an effect on the electricity consumption of freezers and refrigeration. Illumination affects the lighting load and precipitation has the effect of keeping people indoors and so increasing the demand for electricity.

3. FACTORS TO BE TAKEN INTO ACCOUNT IN THE MODEL DESIGN

The following factors were taken into account in the development of a model for the relationship between demand and weather.

- 1) Time-of-use: The relationship between electricity and demand will vary according to the time-of-day and day-of-the-week. On weekdays, the effect of the weather will be particularly noticeable around the peaks, as people arrive home from work. This is more pronounced during winter as it coincides with the lighting -up peak. At

the weekends, the demand is spread more evenly throughout the day.

- 2) Seasonalities: Electricity demand varies considerably across seasons. Lighting loads shift with number of daylight hours. The space-heating load disappears, and is replaced by an insignificant cooling load.
- 3) Nonlinearly: The relationship between the weather variables and demand is non-linear
- 4) Demand growth: The load changes over time as a result of the economic growth, efficiency increases and environmental factors.

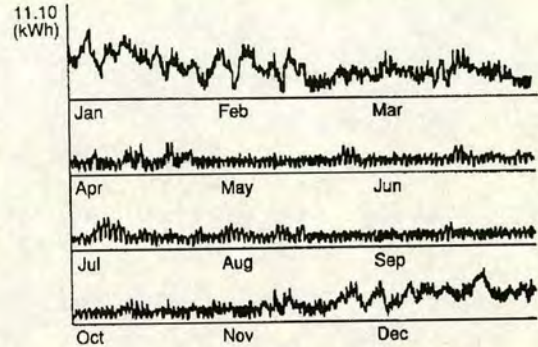


FIGURE 1. Estimated load shape for base household

Figure 1 shows the base household demand for customers of the Boston Edison Company, in 1981 (4). From the figure, it may be seen that household demand is particularly volatile, during the winter months; this is thought to be due mainly to weather effects.

4. MODELS

This section reviews the different options for assessing the relationship between weather and domestic demand.

4.1 Multiple Linear Regression (MLR)

This is the most commonly used method in practice because of its ease of application and flexibility. MLR is used either as a stand-alone method for load forecasting or to define the transfer function in a time-series load forecasting model.

There are several MLR methods for determining the relationship between weather and demand. The method described in this paper is based on reference 3 and uses the degree-day concept.

HDD = heat degree day

CDD = cooling degree day

A threshold temperature is chosen above which it is assumed there is a negligible effect on demand.

HDD= threshold temp - actual temp

CDD= actual temp - threshold temp

For simplicity only the relationship of weather to temperature is defined. In the UK there is a minimal cooling load, so the cooling load is

$$D = a + bHDD + c + dHDD * NBD + \sum_{n=1}^{12} \frac{e_n S_n}{n} + \gamma$$

a=base demand

b=business day demand

c=%change in demand growth

NBD=1 if non business day; 0 if business day

b+d=weekend demand

$e_1 \dots e_{12}$ =monthly dummy variables

$S_1 \dots S_{12}$ =monthly effects

γ = error function

To represent better the non-linear relationship, two regression models can be used with different threshold temperature (4). A MLR can be performed with most statistical packages. The main drawback of this method is that it does not represent anomalous events. For example unusually cold spells, to which companies are most vulnerable.

4.2. Polynomial regression model

Polynomial models can represent higher order functions in the relationship between dependent and independent variable. Given that the relationship between temperature and electricity demand is not linear, a polynomial model may provide a significant improvement over a linear relationship in terms of goodness of fit (5).

Unfortunately, polynomial functions may provide unreliable descriptions at either end of the temperature range (cold spells). This is because the shape of the polynomial is more heavily influenced by data at the dense part of the temperature distribution close to the mean. Hence this method cannot best represent anomalous weather.

4.3 Nonparametric regression

This method was not used much in the past because of the lack of adequate computing power. Nonparametric regression allows the weather-demand relationship to be calculated directly from historical data. The properties of

the relationship are described in terms of a multivariate probability density function (PDF) of demand, time, and weather variables.

Nonparametric methods are highly graphical. For the specific application required, they provide an accurate representation of temperatures at the edges of the range. This method represents anomalous weather events. The non-parametric kernel smoothing regression is generally used, because of its simplicity (6).

4.4 Neural network with fuzzy logic

Thus far we have discussed econometric methods. The method, which has attracted increased interest over the past decade, is Artificial Neural Networks (ANN). The basic principles of an ANN are that it learns patterns from the input, weather variables and output, demand, it then creates its own non-linear model. With the addition of a fuzzy classifier, the different classes of ANN can be used.

The ANN input data is subdivided into different classes, by the fuzzy set, based on the weather conditions. Figure 1. Shows the flow diagram of fuzzy set data classification where N represents the total number of weather condition classes.

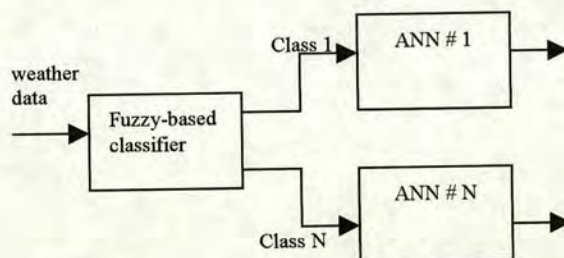


FIGURE 1,Flow diagram of the proposed technique

The classifications are done based on the season, day-of-the-week, and public holidays. The temperature, wind-speed and precipitation are selected as criteria for historical data classification.

Temperature is sorted into eight categories and labelled as Extremely Cold (ExC), Very Cold (VC), Cold (C), Normal (N), Warm (W), Hot (H), Very Hot (VH) and Extremely Hot (ExH). Wind-speed is sorted into four categories, Extremely Windy (ExW), Very Windy (VW),

Windy (W) and Normal (N). Precipitation is sorted into two classes, Wet (W) and dry (D). The categories of temperature and wind speed are then used to form 64 possible classes of weather, 32 for wet, and another 32 for dry weather conditions. The numbers 1 to 64 are assigned to the 64 classes.

The ANN will be assigned to each data class, in order to learn from historical data, the effect of the weather condition on the demand. This method will adequately represent each weather condition. However, the wrong number of hidden neurons would result in under or over learning. This would affect the accuracy of the prediction.

5. DATA REQUIREMENTS

The data requirements of this exercise are extensive, because it is trying to extrapolate accurate relationships from historical data. Both electricity demand and weather variable data are required.

5.1 Location

The specified city for this investigation is Aberdeen, since robust weather and electricity demand data are readily available for the analysis.

5.2 Data specification

Average daily demand as well as peak daily demand will be used. Daily maximum and minimum temperatures for this time period are compiled. The average and maximum daily wind speeds together with precipitation data are also collected.

5.3 Time horizon

A time horizon of six years has been chosen, as it is long enough to depict all the possible trends, but short enough to exclude obsolete trends.

6. TEST PROCEDURE

The following test procedures will be used to compare the performance of the outlined methods.

6.1 Historical

The models will be tested on the most recent year's historic weather data. The outputs will be compared to the actual demand.

6.2 Scenario Analysis

Various scenarios will be simulated and the models performances compared.

6.3 Monte-Carlo

Synthetic weather data will be simulated. The performance of the various models will be compared with the inputted data.

7. CONCLUSION

Increased competition in the Electricity industry has resulted in tighter margins for participants. Understanding the relationship between the fundamental demand driver -weather- and electricity demand is crucial. Three standard econometric techniques and a fuzzy neural network approach are introduced to solve the relationship. The implementation of these models is currently underway, strengths and weaknesses of each approach will be compared and reported in future papers.

REFERENCES

- 1) Green, R.J. and Newbury, D.M., *The Electricity Contract Market*, Journal of Industrial Economics, Forthcoming, 1998.
- 2) Smith Peter, *Personal Communications*, Eastern Energy, Australia, Oct 1998.
- 3) Meed, Carl. *Personal Communications*, Scottish and Southern Energy Plc, Perth, March 1999
- 4) Gellings, P.E., *Demand Forecasting for Electric Utilities*, The Fairmont Press, Inc, pp 41-75, 123-147, 1992. Charytoniuk, W.,
- 5) Peirson, J., Henley, A., *Non-Linearities in Electricity Demand and Temperature; Parametric versus Non-Parametric Methods*, Energy Economics vol., pp149-162.
- 6) Chen, M.S. Van Olinda, P. *Nonparametric Regression Based Short-Term Load Forecasting*, IEEE transactions on Power Systems, Vol. 13, No. 3, Aug 1998
- 7) Daneshdoost, M. Lotfalian, M. Bumroongit G., *Neural Network with Fuzzy Set-Based Classification For Short-Term Load Forecasting*, IEEE transactions on Power Systems, Vol. 13, No. 4, Nov 1998.
- 8) Birch A.P., Smith, M.M. and Ozveren, C.S., *Small area Load Forecasting using Neural Network Approach*, 27th UPEC '92, 23-25 Sept 1992, p181-183.
- 9) Davis, M. *The Relationship between Weather and Electricity Demand*, IEE Monograph No.314S, 1958.
- 10) Gross, G., Galiana, F.D., *Short-Term Load Forecasting*, proceedings of the IEEE, Vol. 75, No. 12, Dec 1987, pp 1558-1572.

AUTHOR'S ADDRESS

Energy Systems Group,
Dept of Electronics and Electrical Eng.
The University of Edinburgh,

e-mail : David.Adjepon-Yamoah@ee.ed.ac.uk
Phone : +44 (0)131 650 5584
Fax : +44 (0)131 650 6554

Determining the weather uncertainties in domestic electricity consumption

D. Adjepon-Yamoah*, H.W. Whittington*, A.Birch**

Department of Electrical Engineering, University of Edinburgh*, Scottish and Southern Energy Plc**

The new bilateral markets proposed by NETA (New Electricity Trading Arrangements) will increase income volatility for Suppliers, because of the replacement of the centralized 'clearing system', the Pool, with bilateral auction-based markets. Added to this, stiffer penalties for out-of-balance market participants will increase supplier volumetric risk. With no historical NETA market data and no singular market price, managing the volumetric exposure to the fundamental electricity demand drivers is imperative. For this reason as complete as possible an understanding of the effects of market fundamentals on demand is required.

One identifiable component of volumetric risk is that of weather dependence. This paper compares two standard models for determining weather uncertainties.

1 Introduction

The introduction of NETA in Autumn 2000, followed by the STA (Scottish Trading Arrangements) due in the next few years will pose a new set of increasingly complex challenges to participants in the Electricity Supply Industry (ESI).

The bilateral markets of NETA will have the following implications for Suppliers

- Higher volatility
- Higher contract prices
- Greater role in the price setting mechanism
- Higher imbalance charges
- No single reference price

The above reasons will lead to greater uncertainty for Suppliers [6]. This in turn, will increase the need for much better understanding of customer habits, to be used in managing demand-side volumetric risk. From such studies should emerge more sophisticated risk management strategies.

1.1 Customer demand

For purposes of this paper, customers are divided into

1. Industrial customers
2. Domestic/commercial customers

1.1.1 Industrial Customers

Since the bulk of the industrial customer load is made up of industrial processes, the demand curve tends to follow similar daily patterns. Excluding anomalies, this demand segment does not cause much uncertainty for suppliers, because it is not very variable.

1.1.2 Domestic/Commercial

Customer consumption is much more variable. For example, on a given day no two households will have identical demand, and even within a given household consumption patterns vary from day-to-day causing much uncertainty for suppliers. For this reason this paper uses the domestic demand segment as the focus for the analysis.

1.2 Fundamental drivers

The fundamental drivers of electricity demand include

- Economic trends
- Time factors (e.g. weekly, seasonal, holidays)
- Weather factors
- Random effects

Experience shows that weather is the largest single driver of electricity demand [8] (see Figure.1). It follows that an understanding of the relationship between weather and demand will help Suppliers predict their long- and short-term exposure to the weather.

1.3 Weather-demand models

In this paper, the two common but generically different models for determining the relationship/effect of the weather on demand are compared.

They are:-

- Multiple Linear Regression
- Artificial Neural Networks

2 Effects of the weather on domestic demand and other related issues

This section describes the weather sensitive components of domestic demand, and outlines the weather factors which directly influence demand. Also covered are related variables which have a bearing on the weather-demand relationship.

2.1 Weather sensitive components of domestic demand

The major weather dependent uses of electricity in domestic households are identified as

1. Space heating
2. Water heating
3. Refrigeration/freezer
4. Others (e.g. kettles, audio visual etc)

Of the above factors, space heating makes up over 75% of the weather sensitive component of demand [10].

2.2 Weather factors affecting demand

The weather effects found to affect demand are:

1. Temperature
2. Wind speed
3. Precipitation

The temperature and wind speed affect the space and water heating demand. During winter or when the weather is bad, all three weather variables have the effect of keeping people indoors, increasing their overall consumption

Figure.1 below shows the correlation between demand and temperature. The analysis was performed on data from a location in the North-east of Scotland. It shows broadly, an inverse relationship between the temperature and demand. There are two temperature-demand clusters, the denser cluster is the weekday demand, and the other is the weekend demand. Though both clusters have roughly the same shape, the graph shows that the average daily demand is less at weekends.

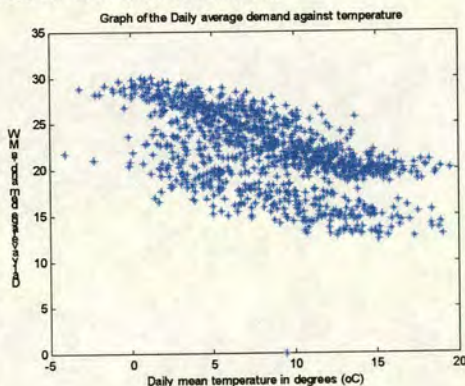


Fig.1 daily average demand v temp

2.3 Other related factors which affect the weather-demand relationship

On a local level, the weather-demand relationship is also directly influenced by the following factors:

- Desired level of thermal comfort
- Duration of cold/hot spell
- Rate of heat build-up (dependent on the density of buildings)

For example, a two bedroom house in the country will be more exposed to the natural elements than a similar same sized property in a densely built up area, resulting in a stronger weather-demand relationship.

3 Data requirements and practical considerations

The data requirements for both weather and electricity demand are extensive if a reasonably accurate model is to be developed.

3.1 Data specification

For electricity demand, the daily half-hourly demand from a 33kV Bulk Supply Point (BSP) supplying a predominantly domestic load was used. Hourly temperatures together with wind speeds and direction were obtained from the Met. Office's nearest weather station to the BSP.

3.2 Time horizon

The time-horizon chosen for this study was three and a half years to ensure robust and continuous data. We considered that this time-period is long enough to depict most possible trends and short enough to exclude most extreme events.

3.3 Active demand management

For the chosen location the first-tier electricity Supplier had installed active Demand Side Management (DSM) to fit better the space and water heating loads to the generating profile [3]. The DSM methods used are tele- and time-switching and a combination of both systems. The daily switch-on and off times for the time-switched customers is fixed but the teleswitched customer load varies according to the forecast minimum day-ahead temperature. This dynamic load management alters the daily half-hourly sensitivities to the weather by shifting the

heating load across half-hours according to the forecast temperature.

3.4 Influence of teleswitched data

The use of teleswitching dynamically affects demand by introducing a bias into the demand-weather relationship.

To dampen the effect of this bias an algorithm was developed which removed the base effect of the teleswitch load, leaving the weather sensitive component of the load. To determine the base teleswitched load, real-time data is used. The base effect is calculated as the mean of the teleswitched demand block on days with a minimum temperature greater than 11 degrees over the autumn/winter period. Because teleswitching is used mainly during the winter months, the days were selected from the months October till April.

When applying the algorithm, we were aware that the number of customers on the different tariffs is very variable. This means that extrapolations may only be valid for a few months, but it was considered a worthwhile exercise, despite this reservation.

4 Model philosophy and construction

This section summarizes the reasons for the model choice and discusses the model construction.

4.1 Model choice

Traditionally various classes of models have been used to forecast electricity; these are mainly grouped into Time-series, Regression based and Artificial Intelligence (AI) models. For extrapolation the latter two are most commonly used [9]. The regression-based methods cover the classes

- Multiple Linear regression
- Polynomial regression
- Non-parametric regression

The AI models include

- Artificial Neural Network (ANN),
- Fuzzy Neural Networks (FNN)

In this paper we compare the two generic classes. A multiple linear regression was used because of the following:-

- It is the most commonly used classical model

- it is also very flexible and easily implemented.

The Artificial Neural Network (ANN) was chosen as the AI method because it is

- The most widely used AI method in demand forecasting.

The ANN performs a non-parametric regression on the data. The model choices could be effectively viewed as a comparison between a linear, parametric method and a non-linear, non-parametric method.

4.2 Model Construction

The next two sections will give a brief description of the model design and construction

4.2.1 Data considerations

Several considerations have to be made in the data preparation for the model design. These are outlined below

- Time-of-use The relationship between the demand and the weather will change according to the time of day. It was decided to assign and train a model for each half-hour because this will capture the time-of-day effects better. This method uses 48 models to establish the weather-demand relationship over a day
- Day-of-week Demand varies according to the day of the week. For example, on business days, the domestic demand is concentrated at the morning and evening peaks, during the weekends or for non-business days, the demand is more evenly spread across the day. This is represented in the model design by day-of-the-week dummy variables
- Seasonalities Electricity demand varies considerably across seasons, for example, the heating load disappears during the summer. Using winter and summer dummies in the models accounts for this effect.
- Demand growth Demand changes with time, for various reasons, lifestyle changes, economic trends etc. For this reason an incremental demand coefficient was considered.

4.2.2 Multiple linear regression

The MLR used in this project is defined below.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 \bar{x} + \beta_3 z_t + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} + \rho_4 y_{t-4} + \sum_{n=1}^6 D_n x_t + \varepsilon_t$$

where

y = demand

x=temperature

\bar{x} =mean lagged 24-hour temperature

z=wind speed

β_0 =the base demand

β_1 =the temperature coefficient

β_2 =the mean of the previous 24 hour temp

β_3 =the wind-speed coefficient

ρ_{1-4} = the lagged previous 2 hour demand coefficients

D_n =the dummy variables for six day-of-week. The Sunday dummy was left out because it is assumed that Sunday has the highest domestic demand.

ε_t =the error term

To perform a forecast, the regression coefficients are obtained by applying the equation to the training dataset and these coefficients are used to obtain the forecast demand.

4.2.2 Artificial Neural Network

The ANN is effectively a black box that learns patterns from a given set of inputs and an output. The network used in this project was a Multi-Layered Perceptron(MLP) with a backpropagation algorithm.

In this exercise a three-layered network was used , that is , an input , hidden and an output layer. There were 13 inputs and one output. The optimal number of neurons in the hidden layer varied according to the half-hour period under consideration.

The inputs to the ANN were as follows

- day-of-week, represented as a five digit binary number
- winter/summer dummy
- temperature
- wind-speed
- mean 24 hour lagged temperature

- 2 hour lagged demand

The output to the neural-network was the historical demand.

5 Analysis and discussion of results

The models were constructed and used to forecast on a test data set. For the preliminary analysis, models for two half hours only were built. The half-hours were:

- 00hr-00:30hr
- 12:30hr-13:00hr

The first half-hour was chosen because the time-switches are 'on' during this period, implying the timeswitched load. The second half hour was selected because, on very cold days (<1 degree minimum day-ahead forecast temperature) the teleswitches are 'on' during this period. The two half-hours were consider adequate for the comparison because they reflect the demand-side management used in the area. The training data set had 1095 days and the test set had 90.

5.1 Test method

The analysis could be interpreted as a scenario analysis over a 90 day period on the chosen half-hours. The forecasting accuracy of the technique was evaluated by the average of absolute percentage errors of the half-hours. The absolute percentage error is

$$APE = \frac{|D_{actual} - D_{forecast}|}{D_{actual}} * 100$$

where D_{actual} and $D_{forecast}$ are the actual and forecast demand. The mean absolute percentage error (MAPE) is

$$MAPE = \frac{1}{N_h} \sum_{N_h} APE$$

where N_h is the number of half-hours in the forecast period.

5.2 Results

The MAPE for the MLR and the ANN for both half-hours were calculated and presented in table1.

Table1. Forecasting results in terms of MAPE

MAPE %	00hr-00:30hr	12:30-13:00hr
MLR	2.8442	4.0101
ANN	1.8630	1.0314

The results in table 1 indicate that the ANN outperforms the MLR for both half-hour periods. While the ANN performs better for the second half-hour than the first, the MLR gives worse results for this half-hour. This is the case because the teleswitch introduces greater non-linearity into the relationship. As the MLR is a linear approximation of the relationship, the errors increase.

To determine the performance of the models at different temperatures, the half-hour 00hr-00:30hr was used. This is because it does not have the added non-linearity induced by the teleswitch.

Table 2 shows the MAPE for both models for the temperature ranges:-

- $t \leq 0$ degrees
- $0 \leq t \leq 5$ degrees
- $t > 5$ degrees

Table 2. MAPE of the models for temperature ranges

MAPE %	$t \leq 0$	$0 \leq t \leq 5$	$t > 5$
MLR	5.5151	3.0701	1.4519
ANN	1.9487	1.9438	1.4999

From table2, the ANN performs much better than the MLR at lower temperatures. For temperatures above 5 degrees, the performance of both models is very similar. Figure 2 illustrates a similar trend by plotting the %APE of both models against the temperature.

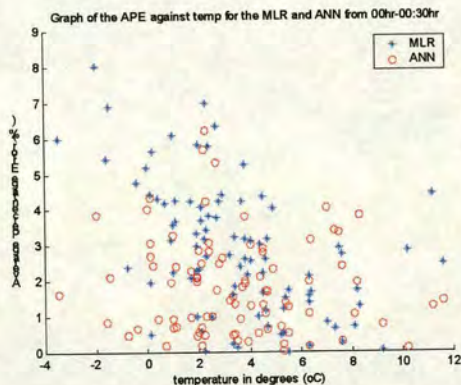


Fig 2. %APE v temperature for both models

5.3 Discussion

From the analysis in the previous section, the ANN outperforms the MLR in all the categories assessed. This is because the weather-demand relationship is a non-linear relationship and the

ANN gives a better approximation of non-linearity than the MLR. Also the ANN is able to account for the additional non-linearity caused by the teleswitch. The MLR has several advantages over the ANN. These are listed below:-

- Easier implementation and greater flexibility.
- Model coefficients are visible enabling easier inference

6 Application of the models

This final section gives a brief summary of the short to long term applicability of these models in Suppliers risk management activities. Firstly the issue of accuracy and calibration are considered.

6.1 Accuracy and calibration

Over time the accuracy of the chosen model will decrease because, the weather demand relationship is dynamic. These changes are caused by appliance efficiency increases, customers changing heating fuel source and demand growth. The chosen model has to be re-trained fairly regularly to maintain a reasonable accuracy

6.2 Application

- Short-term:- The model could be used to determine a suppliers day-ahead exposure to weather, given forecasts. From this the volumetric imbalance caused by the weather is obtained.
- Medium-term:- A supplier could calculate its weather exposure for a period of over a few months for risk management purposes.

7 Conclusion

Increased competition in the Electricity industry has resulted in tighter margins for participants. Understanding the relationship between the one important domestic fundamental demand driver – weather- and electricity is crucial. Two generically different models for determining the relationship – Multiple Linear Regressions (MLR) and Artificial Neural Networks (ANN) – were compared. Increased non-linearity had been introduced into the data set by active Demand-Side Management (DSM) in the area.

From the analysis, the ANN outperforms the MLR. This is because the ANN is capable of modelling the non-linearity induced by the DSM and that in the weather-demand relationship. The MLR has the advantages of transparency and easy implementation.

In choosing the best model for practical application, not just performance, but all other factors have to be considered.

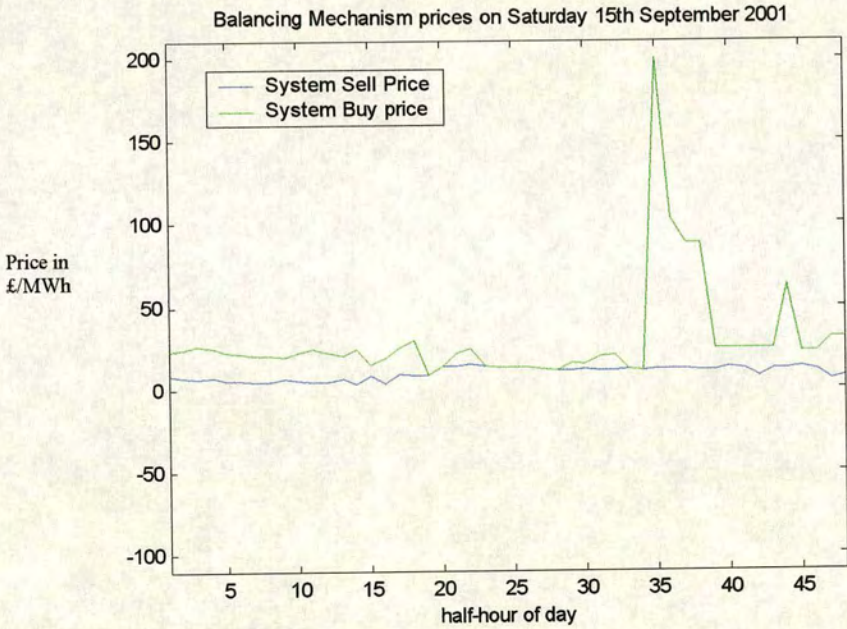
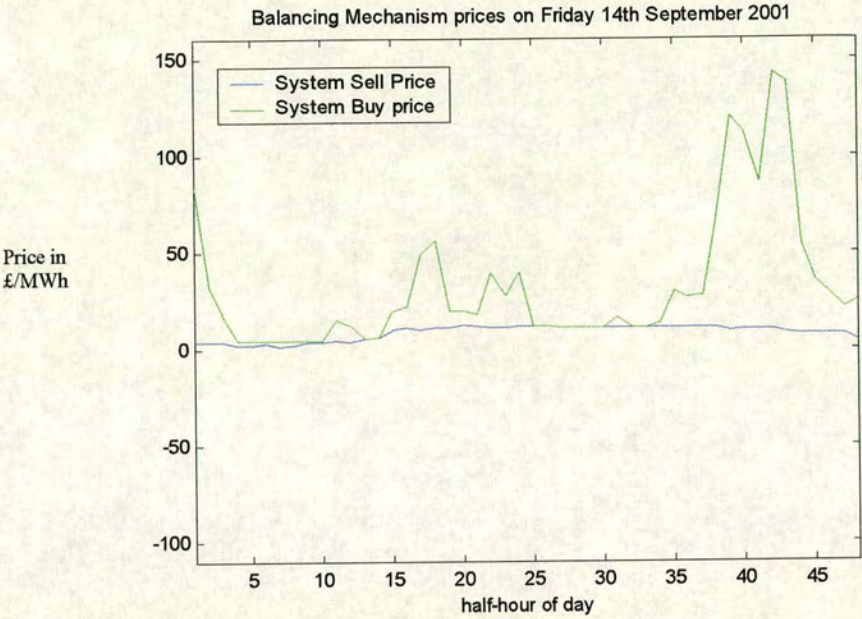
REFERENCES

- 11) Anderson B. R., Chapman P. F., Cutland N. G., Dickson C.M., Shorrock L.D., *BREDEM-12 : Model description*. Building Research Establishment Laboratory Report, 1996.
- 12) Birch A.P., Smith, M.M. and Ozveren, C.S., *Small area Load Forecasting using Neural Network Approach*, 27th UPEC '92, 23-25 Sept 1992, p181-183.
- 13) Connolly, John. *Personal Communications*, Scottish and Southern Energy Plc, Pitlochry, September 1999
- 14) Daneshdoost, M. Lotfalian, M. Bumroongit G., *Neural Network with Fuzzy Set-Based Classification For Short-Term Load Forecasting*, IEEE transactions on Power Systems, Vol. 13, No. 4, Nov 1998.
- 15) Davis, M. *The Relationship between Weather and Electricity Demand*, IEE Monograph No.314S, 1958.
- 16) *Electricity Trading Seminar*, IEE 00/038, March 2000
- 17) Freud, F.J., Wilson, W.J., *Regression Analysis: Statistical Modeling of a Response Variable*, Academic press, 1998
- 18) Gellings, P.E., *Demand Forecasting for Electric Utilities*, The Fairmont Press, Inc, pp 41-75, 123-147, 1992. Charytoniuk, W.,
- 19) Gross, G., Galiana, F.D., *Short-Term Load Forecasting*, proceedings of the IEEE, Vol. 75, No. 12, Dec 1987, pp. 1558-1572
- 20) Peirson, J., Henley, A., *Non-Linearities in Electricity Demand and Temperature; Parametric versus Non-Parametric Methods*, Energy Economics vol., pp149-162.

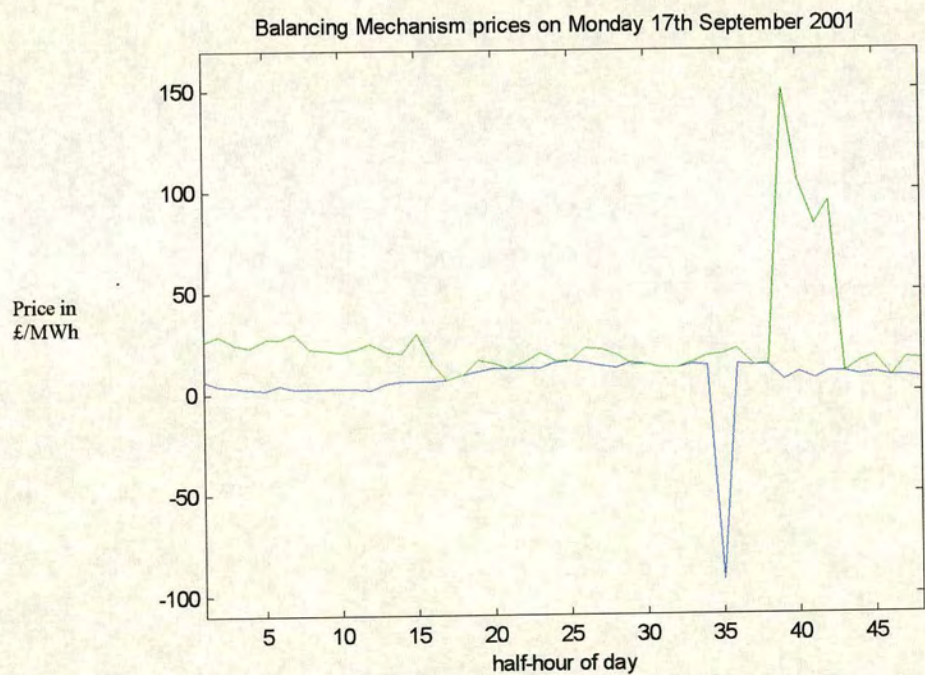
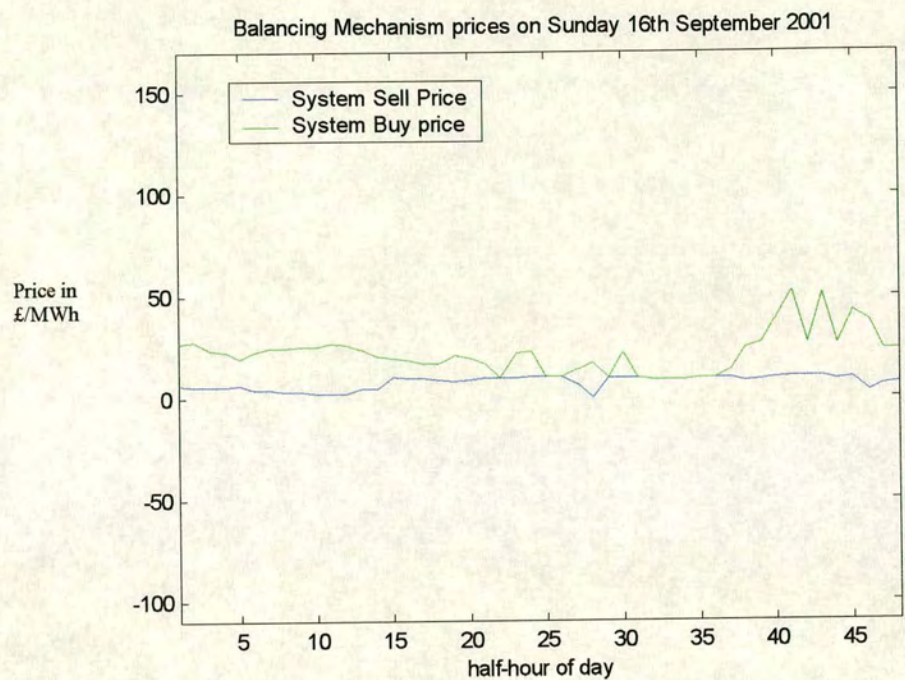
Appendix 2

**Half-hourly Balancing Mechanism System Buy Price and System
Sell Price from 14th – 20th September 2001**

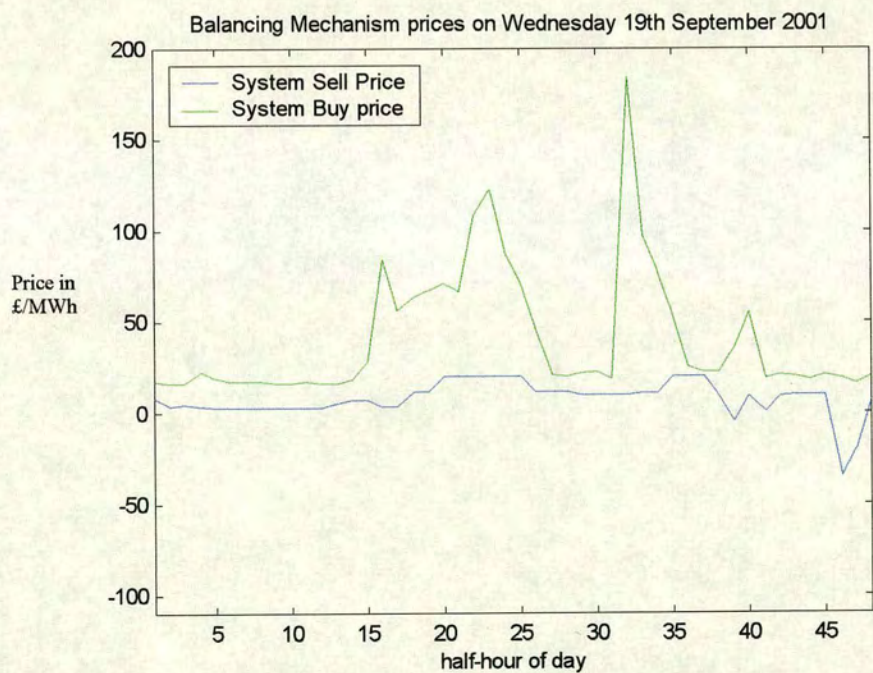
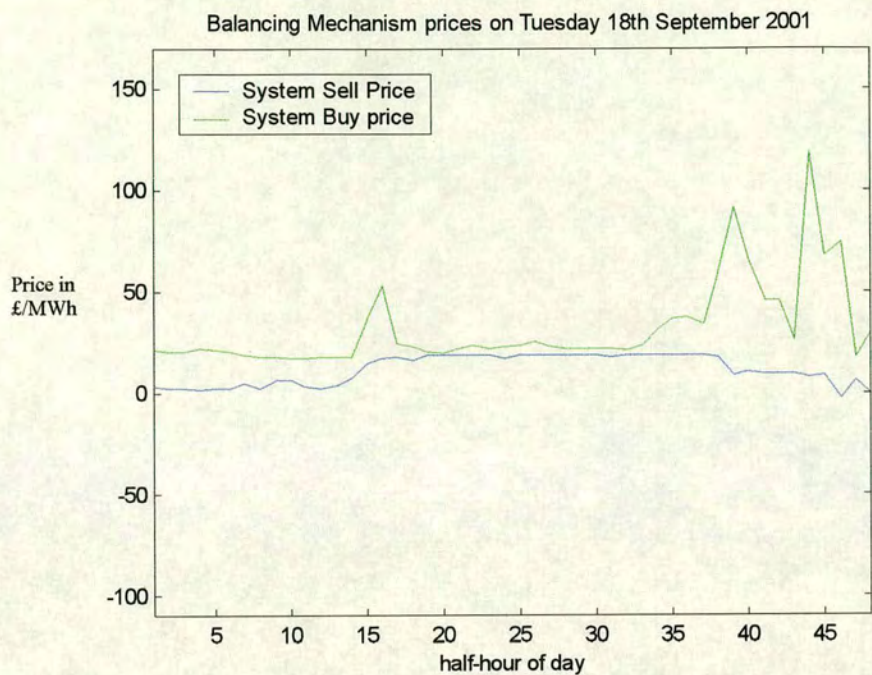
Appendix 2



Appendix 2



Appendix 2



Appendix 2

