Dynamic Modelling of Generation Capacity Investment in Electricity Markets with High Wind Penetration

Daniel Eager

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
2012

Abstract

The ability of liberalised electricity markets to trigger investment in the generation capacity required to maintain an acceptable level of security of supply risk has been - and will continue to be - a topic of much debate. Like many capital intensive industries, generation investment suffers from long lead and construction times, lumpiness of capacity change and high uncertainty. As a result, the 'boom-and-bust' investment cycle phenomenon, characterised by overcapacity and low prices, followed by power shortages and high prices, is a prominent feature in the debate. Modelling the dynamics of generation investment in market environments can provide insights into the complexities involved and address the challenges of market design.

Further, many governments who preside over liberalised energy markets are developing policies aimed at promoting investment in renewable generation. Of particular interest is the mix and amount of generation investment over time in response to policies promoting high penetrations of variable output renewable power such as wind. Consequently, improved methods to calculate expected output, costs and revenue of thermal generation subject to varying load and random independent thermal outages in a power system with a high wind penetration are needed.

In this interdisciplinary project engineering tools are applied to an economic problem together with knowledge from numerous other disciplines. A dynamic simulation model of the aggregated Great Britain (GB) generation investment market has been developed. Investment is viewed as a negative feedback control mechanism with current and future energy prices acting as the feedback signal. Other disciplines called upon include the use of stochastic processes to address uncertainties such as future fuel prices, and economic theory to gain insights into investor behaviour. An 'energy-only' market setting is used where generation companies use a classical NPV approach together with the Value at Risk criterion for investment decisions. Market price mark-ups due to market power are also accounted for.

The model's ability to simulate the market trends witnessed in GB since early 2001 is scrutinised with encouraging findings reported. A reasonably good agreement of the model with reality, gives a degree of confidence in the realism of future projections. An advancement to the dynamic model to account for expected high wind penetrations is also included. Building on the initial model iteration, the short-term energy market is simulated using probabilistic production costing based on the Mix of Normals distribution technique with a residual load calculation (load net of wind output). Wind speed measurement data is combined with the outputs of atmospheric models to assess the availability of the GB wind resource and its relationship with aggregate load.

Simulation results for 2010-40 suggest that the GB system may experience increased generation adequacy risk during the mid to late the 2020s. In addition, many new investments are unable to recover their fixed costs. This triggered an investigation into the design of a capacity mechanism within the context of the modelling environment. In light of the ongoing GB market electricity market reform debate, two mechanisms are tested; a strategic reserve tender and a market-wide capacity market. The goal of these mechanisms is to mitigate generation adequacy risk concerns by achieving a target winter peak de-rated capacity margin.

Declaration of originality

I hereby declare that this thesis was composed by myself and that, except where indicated, the research recorded is entirely my own.

Dan Eager

Acknowledgements

I would like to thank Janusz Bialek for his sound advice and support over the duration of this research. Thanks to Tim Johnson for providing useful insights, encouragement and reassurance. This was particularly helpful during the final stages. I would also like to Gareth Harrison, whose door was always open, for his understanding and direction throughout the final stages.

A special thanks to Ben Hobbs, for his significant input into the project and help developing many of the novel elements presented. His enthusiastic and patient guidance before, during and after my placement with EPRG was vital in pushing my research forward. My thanks to EPRG, in particular David Newbery, for being fantastic hosts and for facilitating the collaboration.

I would also like to acknowledge the valuable input from Chris Dent who has helped with so many elements of the work. Thanks to Colin Gibson who allowed me access to his wealth of industry knowledge. Many thanks to Jim Skea for sharing passing on his data. Also thanks to Sam Hawkins for the use of his model outputs and for taking the time to help me make sense of it all.

Many thanks to the numerous people at National Grid; the knowledge transferred during my placement has proved invaluable.

A huge thanks to my family and friends for their advice, but most importantly the social distractions. I am grateful to Mum, Dad and Adam for their loving support, which has been vital for everything I have achieved. I would especially like to thank my girlfriend Kate who has endured the stresses and strains of this process more so than anyone else. She has been a relentless source of encouragement and has listened patiently and lovingly to my complaining throughout.

My thanks go to my friends at in the Institute for Energy Systems. In particular, those who were unfortunate enough to share an office with me; the relentless jovial atmosphere has kept me sane. Special mention to Adam, Anup, Barry, Brian, Camilla, Chaz, Dougal, Evangelos, James, Jorge, Nacho, Paddy, Punim, Randy, Richard, Ronan, Rudra, Sarah and Sorin.

Many thanks to my friends in the UKERC Sparks network (too many to list) whose encouragement and social banter before, during and after our countless research symposiums was vital for maintaining moral.

The award of a studentship from UKERC for the duration of study is gratefully acknowledged as providing me with a means of support during this process. This research formed part of the UKERC programme and was supported by the UK Research Councils under NERC award NE/C513169/1. Also, the award from the Royal Academy of Engineering for financing my trip to the 2010 IEEE PES General Meeting is gratefully acknowledged.

Abbreviations

ACS	-	Average cold spell
ACSWP	-	ACS winter peak
ASC	-	Advanced supercritical
AEB	-	Area electricity boards
AS	-	Ancillary services
IFA	-	Anglo-French Interconnector
BSC	-	Balancing and Settlement Code
BM	-	Balancing Mechanism
BETTA	-	British Electricity Trading and Transmission Arrangements
BERR	-	Business Enterprise & Regulatory Reform, Department for
CC	-	Capacity credit
CF	-	Capacity factor
CM	-	Capacity margin
COPT	-	Capacity outage probability table
CRF	-	Capital recovery factor
CCS	-	Carbon capture and storage
CO_2	-	Carbon Dioxide
CEGB	-	Central Electricity Generating Board
CLP	-	Closed loop policy
CCGT	-	Combined cycle gas turbine
CHP	-	Combined heat and power
CCC	-	Committee on Climate Change
CR	-	Concentration ratio
CVaR	-	Conditional value at risk
CfDs	-	Contracts for differences
CEFF	-	Convert, enrich and fuel fabricate
cdf	-	Cumulative distribution function
DDE	-	Delay Differential Equation
DR	-	Demand Response
DECC	-	Department of Energy and Climate Change
DRC	-	De-rated capacity
DRM	-	De-rated capacity margin
DRW	-	De-rated wind capacity
DHHI	-	Dynamic Herfindahl-Hirschman Index
ESDS	-	Economic and Social Data service
ELCC	-	Effective load carrying capability
EFC	-	Effective firm capacity
ESP	-	Electricity service provider
ESI	-	Electricity supply industry
ETS	-	Emissions Trading Scheme
E&W	-	England and Wales

EU	-	European Union
EEU	-	Expected energy unserved
EEU	-	Extra required capacity
FAFC	-	Full amortised fixed cost
FiT	-	Feed-in-tariff
FPN	-	Final physical notification
FC	-	Fixed cost
FGD	-	Flue-gas desulpurisation
FDRM	-	Forecast de-rated margin
FOR	-	Forced outage rate
FAFC	-	Full amortised fixed cost
GT	-	Gas turbine
GQ	-	Gaussian quadrature
GW	-	Giga-watt
GWEA	-	Global Wind Energy Agency
GB	-	Great Britain
GHG	-	Greenhouse gas
GM	-	Gross margin
HHI	-	Herfindahl-Hirschman Index
IPP	-	Independent power producer
ISO	-	Independent system operator
ICAP	-	Installed Capacity
IRM	-	Installed reserve margin
IGCC	-	Integrated gasification combined cycle
IPCD	-	Integrated prevention control directive
IRR	-	Internal Rate of Return
kW	-	Kilo-watt
kWy	-	Kilo-watt-year
LCPD	-	Large combustion plant directive
LI	-	Lerner Index
LC	-	Levelised cost
LLD	-	Load limiting device
LSI	-	Load serving entity
LDC	-	Load-duration curve
LMP	-	Locational marginal pricing
LRMC	-	Long-run marginal costs
LOLE	-	Loss-of-load expectation
LOLP	-	Loss-of-load probability
MIP	-	Market Index Price
MWd	-	Mega-watt-day
MWh	-	Mega-watt-hour
MBTu	-	Million British thermal units
MOND	-	Mix of Normals distribution
MC	-	Monte Carlo
MPPD	-	Most probable peak demand
NG	-	National Grid
NGC	-	National Grid Company

NETA - New Electricity Trading Arrangements
NPV - Net present value
NO_x - Nitrogen oxide
NSHEB - North of Scotland Hydro-Electric Board
NE - Nuclear Electric
OFGEM - Office for Gas and Electricity Markets
OCGT - Open cycle gas turbine
OTC - Over-the-counter
OC - Overnight cost
PJM - Pennsylvania-New Jersey-Maryland
PSI - Pivotal supply indicator
PPP - Pool purchase price
PSP - Pool selling price
PCMI - Price-cost margin index
PDC - Price-duration curve
pdf - Probability density function
pmf - Probability mass function
PI - Profitability index
PS - Pumped storage
Q - Quantity
RPM - Reliability Pricing Model
RO - Renewables Obligation
ROC - Renewables Obligation Certificate
RSI - Residual supply index
RRH - Run-of-river hydro
SSE - Scottish and Southern Energy
SP - Scottish Power
SRMC - Short-run marginal cost
STOR - Short-term operating reserve
SSEB - South of Scotland Electricity Board
SO_2 - Sulphur dioxide
SFE - Supply function equilibrium
SMA - Supply margin assessment
SD - System Dynamics
SBP - System buy price
SMC - System marginal cost
SMP - System marginal price
SO - System operator
SMA - Supply margin assessment
SSP - System sell price
TAFC - Total annualised fixed costs per unforced MW
TAE - Total available energy
TIC - Total installed capacity
TIAC - Total interest accumulated during construction
TGC - Tradable green certificate
UCAP - Unforced capacity
UEC - Unit effective capacity

UK	-	United Kingdom
US	-	United States
VaR	-	Value at risk
VRR	-	Variable Resource Requirement
VOLL	-	Value-of-lost-load
VC	-	Variable cost
WACC	-	Weighted average cost of capital
WP	-	Wind production
WNA	-	World Nuclear Association

Symbols

α	-	Plant operating lifetime (e.g., years)
β	-	Calibrated exponential constant in aggregate investment response function
χ	-	Gearing ratio (debt/equity)
δ	-	Plant decommissioning time (e.g., years)
ϵ	-	Demand elasticity
γ	-	Expected bond return
η^B	-	Time investment capacity block triggered
η^M	-	Time mothballed capacity block triggered
κ	-	Required investor equity return
λ	-	Speed of mean reversion
μ	-	Mean of Normal distribution
ν	-	Plant thermal efficiency
π	-	Market price
π^*	-	Equilibrium market price
ϑ	-	Carbon produced by burning the fuel at 100% efficiency
ρ	-	Plant forced outage rate
σ	-	Standard deviation of Normal distribution
τ	-	Plant build time (e.g., years)
τ^D	-	Plant de-mothball time (e.g., years)
ω_x	-	Number of capacity blocks under construction
ω_x^M	-	Number of capacity blocks mothballed
ξ^B	-	Investment capacity block
ξ^M	-	Mothball capacity block
ξ_{max}	-	Maximum annual investment block per technology
ψ^B	-	Vector of new build capacity blocks
ψ^M	-	Vector of mothballed capacity blocks
a	-	No load cost (MBTu)
b	-	Average heat rate (MBTu/MWh)
c_u	-	Capacity of unit u
c_x	-	Cost of decommissioning technology x (e.g., £/kW)
d	-	Demand (e.g., GWs)
d^*	-	Normalised and re-scaled demand (e.g., GWs)
h_n	-	Probability generator n is on the margin
i_x	-	Constant used to calibrate aggregate investment response function for technology x
m_u	-	Number of units of type u
$m(t)$	-	Stochastic process drift function
p_c	-	Construction cost (e.g., £/kW))
p_f	-	Fixed operating cost (e.g., £/kW))
p_n^{disp}	-	Probability of dispatch of generator n
q	-	Market quantity
q^*	-	Market equilibrium quantity

$q\%$	-	Critical level in VaR test
$q(t)$	-	DECC fuel price estimate for year t
r	-	Discount rate (or Weighted Average Cost of Capital)
r^i	-	Rate of inflation
r^n	-	Nominal discount rate
r^r	-	Real discount rate
u	-	Discount factor
uf_x^i	-	Plant x utilisation factor, year i
$v(t)$	-	Stochastic process volatility function
$w(L, G_N^*)$	-	Price *mark-up* function
w	-	Weighting coefficient
A_x^{crf}	-	Deferred CRF (for plant type x)
AGM	-	Average gross margin
$C(P)$	-	Variable operating cost
CGM_u	-	Perfectly competitive annual gross margin
D_t	-	Demand in period t
$D^{-1}(q)$	-	Inverse demand function
DC_x	-	Present worth of the decommissioning cost (for plant type x)
F	-	Fuel cost (e.g., £/MWh)
F_{car}	-	Cost of emissions (e.g., £/kg)
F_g	-	Gas price (e.g., £/therm)
F_c	-	Coal price (e.g., £/kg)
F_u	-	Uranium price (e.g., £/kg)
$FCDM$	-	Forecast de-rated margin
G_N	-	Generator n total available capacity from
G_N^*	-	Total available generation
IC_x	-	Investment cost, including interest accumulated during construction (for plant type x)
I_x	-	ICAP (for plant type x)
$I(t)$	-	Total ICAP
L	-	Load
M_i^x	-	Capital expenditure vector (for plant type x)
M_i^q	-	Incidence of capital outlay during decommissioning
M_N	-	Surplus margin
$MR(t)$	-	Stochastic process mean reverting level
N_B	-	Number of base-load plants
N_P	-	Number of peaking plants
OC	-	Overnight (or construction) cost
P	-	Power (MW)
PI_x^q	-	Profitability index (for plant type x)
Q_w	-	Quantity withheld
R_x	-	Set of all retired plant (for plant type x)
RDC	-	Forecast required de-rated capacity
RPM_u	-	Expected revenue from price mark-up
$S^{-1}(q)$	-	Inverse supply function
$U(t)$	-	Utility function
V	-	Variable operation and maintenance cost
V_x	-	Value of an investment (for plant type x)
V_x^{opt}	-	Minimal acceptable V_x (for plant type x)

Contents

List of figures

List of tables

Chapter 1
Introduction

1.1 Background

So far, in many countries, energy market liberalisation has been seen as a success in terms of efficiency and lower prices. Yet worries still remain. One of the main concerns is that the onus is on privately owned generating firms to predict and respond to supply shortages. They must invest in a mix and amount of generation capacity that will maintain an adequate level of security of supply (or generation adequacy) risk. Many governments are seeking to reduce greenhouse gas (GHG) emissions from the power sector by developing policies promoting investment in low carbon and renewable generation. This has introduced uncertainties that liberalised energy markets are struggling to handle efficiently. As the International Energy Agency states: *"Improving decision-making processes for investment related to generation capacity may be one of the most significant benefits of liberalisation. But it is also one of the most serious challenges in liberalised electricity markets"* [1].

Pure 'energy-only' electricity markets are epitomized by payments for supplied energy being the only source of revenue for generators to cover their fixed and variable costs and provide adequate returns on investment. As a result, short-run profits (scarcity rents) are the primary mechanism to mitigate long-term generation adequacy problems. Their ability to trigger investment in the generation capacity required to maintain an acceptable level of security of supply risk is one of the most hotly debated topics in power systems economics. The question is asked: are short-run price signals strong enough to ensure long-term generation adequacy? Further, critics of 'energy-only' markets argue that because the framework is relatively new and constantly changing, it has not been properly proven and the market must be sufficiently concentrated in order to ensure high enough prices to cover fixed costs [2].

Modelling the dynamics of merchant generation investment can further our understanding of the feedback mechanisms that exist between market conditions, capacity investment and subsequent levels of security of supply risk. By engineering dynamic projection models, the debate about whether governments will meet their environmental, security of supply and customer cost goals can be informed. Furthermore, whether the system will experience reliability, price and capacity oscillations in the meantime can be assessed. This is neatly summarised below:

"*Economics focuses on equilibria but has little to say about the dynamics of a market. Once economics shows that a system has a negative feedback loop so that there is a point of balance, it considers its job done. Engineers move beyond this stage of analysis to consider whether a system will sustain oscillations and, if not, whether is is over- or under-damped.*" - Stoft [3].

Further, in a speech delivered to the Royal Society, the governor of the Bank of England said: "*Predicting the precise timing and dynamics of instability ... is very difficult. It is, however, possible to identify a sand pile or forest as being prone to large-scale instability and determine the factors that contribute to that instability ... Economists have been able to learn from other disciplines about how to cope with these types of instability ... [they] have learned much from engineers about how to control dynamic systems.*" [4].

Recently, particularly in Great Britain (GB), interest has turned to the response of the market to policies promoting investment in variable generation such as wind. This has highlighted the limitations of many existing long-term simulation models, and so there is a need to develop new methods of calculating expected output, costs and revenue of thermal generation subject to varying load and random independent thermal outages in a power system with a high wind penetration.

1.2 Project objectives and scope

The project has a number of distinct objectives:

1. To gain an understanding of the complexities of power systems economics, in particular generation capacity investment in market environments, and the methods available to represent these complexities within a simulation environment.

2. To explore the nature of the feedback mechanism that exists between market conditions and generation investment by developing a long-term dynamic generation investment simulation model. This should capture the negative feedback loop that exists between

generation capacity investment and wholesale market prices and how year-on-year investment decisions influence long-term market dynamics.

3. To examine the current status of research concerned with the impact of policies promoting investment in wind generation on thermal plant investments. This includes identifying the limitations of existing approaches, specifically the need to develop new methods of calculating expected output, costs and revenue of thermal generation subject to varying load and random independent thermal outages in a power system with a high wind penetration.

4. To apply the dynamic market model to an existing market, in this case Great Britain (GB). For simplicity, network effects will be neglected and the market will be modelled as a 'single bus' system.

5. To investigate the need for, and design of, capacity mechanisms appropriate for the GB market test case, and explore scope for their implementation within a dynamic modelling environment.

1.3 Thesis and contribution to knowledge

Overall, the research will test the hypothesis that:

> 'energy-only' markets (i.e., without capacity mechanisms) with high penetrations of wind generation are capable of inducing adequate and timely generation investments over a long-term time frame.

The project should provide insight into the nature of capacity margin oscillations in the GB generation market. Further, the likely response of liberalised 'energy-only' markets to policies promoting investment in renewable generation such as wind is not yet fully understood. The work presented here helps move this field of knowledge forward. By extending and applying existing probabilistic production methods to the GB market, the impact of a high wind penetration on investment dynamics can be investigated. As part of the application to a nonequilibrium oligopoly market, a method for estimating revenues from price mark-ups due to market power within the dynamic simulation is derived.

It is envisaged that the modelling techniques, analysis and insights herein will be of interest to not just the UK Government, but all governments seeking to reduce GHG emissions from the power sector whilst maintaining an adequate level of security of supply risk. It would also be of major interest to policy makers, regulators, market designers and others involved in planning and design of the power sector.

1.4 Thesis outline

Chapter 2 provides an overview of the fundamental economics under-pinning modern power markets and the costs and associated investment risks for different generating technologies. A number of existing market frameworks are reviewed with close attention paid to the impact that high standards of system reliability have on market design. In addition, a variety of existing capacity mechanisms are surveyed. Finally, methods to measure generation adequacy risk, including how they are adapted to account for high penetrations of wind generation, are discussed.

Chapter 3 takes a close look at the GB power market. This begins with a brief history of the power industry in GB that illustrates how security of supply concerns have been addressed to date. This is followed by an account of liberalisation in the early 1990s including an overview of market structures. Some policy-informing modelling works applied to the GB market in recent years are also reviewed.

Chapter 4 contains a comprehensive literature survey of the tools available to those modelling generation capacity investment in electricity markets. This includes a review of modelling methodologies, the handling of uncertainty, market power and investor expectations and risk preferences. Finally, a discussion of the scope and value of techniques from control theory to model the investment market is included.

Chapter 5 presents the methodologies used to address the research question posed. To begin, the fundamental model design concepts together with their application to a dynamic investment model of the GB power market are presented. Included are details about all elements of the first iteration of a model. Also included are the results and discussion accompanying this stage of the work. The model's ability to simulate the market trends witnessed in GB since early 2001 is tested along with a discussion of model limitations.

Chapter 6 describes the cutting-edge techniques used to model production from high penetrations of wind power. These results are used to update the investment market simulation model to account for high penetrations of wind. Moreover, the work presented helps answer the following questions: 1) given historic wind resource availability in GB, what would have been the hourly production for a higher level of installed wind capacity; and 2) what is the likely contribution from wind to meeting peak demand?

Chapter 7 has two foci. Firstly, in light of the preliminary results and need for developing computationally fast models, an improved technique for calculating expected output, costs and revenues from the energy market is developed. This includes a method of estimating generator revenues from price mark ups due to market power in an oligopolistic market. The technique uses probabilistic production costing that considers the annual load curve and convolves it with generator outages using the Mix of Normals distribution (MOND) approximation. The accuracy and speed of the technique is tested with encouraging findings reported. Secondly, the integration of the MOND technique in the dynamic simulation model of the GB market is presented.

Chapter 8 explores the scope for including a capacity mechanism within the dynamic model. To begin, the challenges faced when applying techniques from classical control theory to design a robust economic controller are discussed. This motivates consideration of existing modelling methodologies, which given the model application, can provide new insights. In light of the ongoing electricity market reform debate, two mechanisms, namely a strategic reserve tender and a market-wide capacity market are tested. These aim to 1) mitigate generation adequacy risk and 2) make generator revenue streams more predictable.

Finally, Chapter 9 presents headline results, with a discussion of their validity and implications in a GB market context. Also the modelling methodologies and implications to the generation adequacy problem are reviewed. Lastly, final thoughts and thesis experiences are shared.

1.5 How to read this thesis

If the reader is familiar with the fundamentals of power system economics and the generation adequacy problem then Chapter 2 can be overlooked. Similarly, if the history and workings of the GB market are already fully understood, then Chapter 3 can also be skipped, although the historic generation adequacy risk calculation presented in Section 3.5 may be of interest. Further, the literature review of Chapter 4 can be overlooked if the contribution of this thesis, which is presented in Chapters 5 - 8, is of primary interest.

Chapter 2
Electricity Supply and Markets

In order to grasp the complexities of generation capacity investment in liberalised energy markets, it is necessary to understand how the electricity supply industry (ESI) is structured. This chapter provides an overview of the fundamental economics under-pinning modern power markets, such as short-run marginal cost and long-run market equilibrium. Also included is a discussion about the costs and associated investment risks for different generating technologies. A number of existing market frameworks are discussed, and particular attention is paid to the impact that high standards of system reliability have on market design. In addition, a variety of existing capacity mechanisms are reviewed. Finally, methods to measure generation adequacy risk in power systems, including adaption for high penetrations of wind generation, are discussed.

2.1 Introduction

Various structures of ESI exist around the world. Broadly speaking, these are distinguished by the level of *vertical* and *horizontal integration* (or ownership) and *competition* within their component parts; namely the generation, transmission and distribution of electricity within a geographical area (see Chapter 1 of [5]). This work focuses on the generation sector of a liberalised electricity market.

Chile was the first country to liberalise its electricity market in 1986. It was followed by GB in 1990 and Norway in 1991 [6]. Since then, a number of other countries have established wholesale energy markets: Australia (firstly in Victoria, 1994), Brazil (1995), New Zealand (1996), Germany (1998), Spain (1998) and Canada (2002) to name a few. By 1998 many regions of the US also had liberalised ESIs involving Independent System Operators (ISO) for California,[1] Texas (ERCOT), Pennsylvania-New Jersey-Maryland (PJM), New York, New

[1]California was the first US market to move to market-based pricing in 1998.

England and the Midwest. Over the years, each liberalised market has evolved differently. The intimate workings of each are beyond the scope of this study, though a wealth of information can be found in [6].

Although market design differs between countries, the goal of all liberalised electricity market designers is to ensure that the industry operates in the best interests of society. This includes providing affordable, secure and reliable electricity for consumers. They must also be designed in order to remain liquid, promote competition and reduce barriers to entry wherever possible.

2.2 Overview of wholesale energy markets

In the presence of *wholesale competition*, generating firms (or generators) trade directly with large industrial loads and retailers acting on behalf of consumers. Generators can be vertically integrated or independent power producers (IPPs). The generation market structure varies, with oligopoly being the most common structure seen in liberalised markets (e.g., GB, Germany, Spain, Italy). An oligopoly market is a situation where a relatively small number of competing firms dominate. The market is deemed 'energy-only' if payments for energy are the primary source of revenue for generators.[2] Examples of 'energy-only' markets currently operating include GB, Australia's National Electricity Market, Alberta, Nordpool, Ontario, and Texas.

Transactions take place on the *wholesale market* (Fig. 2.1). In this market, the commodity is energy, typically, a mega-watt hour (MWh), which is the flow of power (MW) over a period of time (h) for an agreed price (£/MWh). The process of committing (or dispatching) generating units in order to meet demand is called *unit commitment*. Sometimes a distinction is made between *unit commitment* and *energy dispatch*. More precisely, unit commitment pertains to unit start and stop times, and level of generation, whereas energy dispatch relates to the level of generation only. Depending on market structure, this is either operated by the System Operator (SO) or involves private (or self) dispatch independent of the SO. In general terms, unit commitment simply informs generators when to switch on and how much to produce [3].

To date, it remains uneconomic to store electricity in large quantities, therefore electricity gen-

[2] Some generators can obtain additional revenues by providing ancillary services (AS); this is discussed in Section 2.2.3.

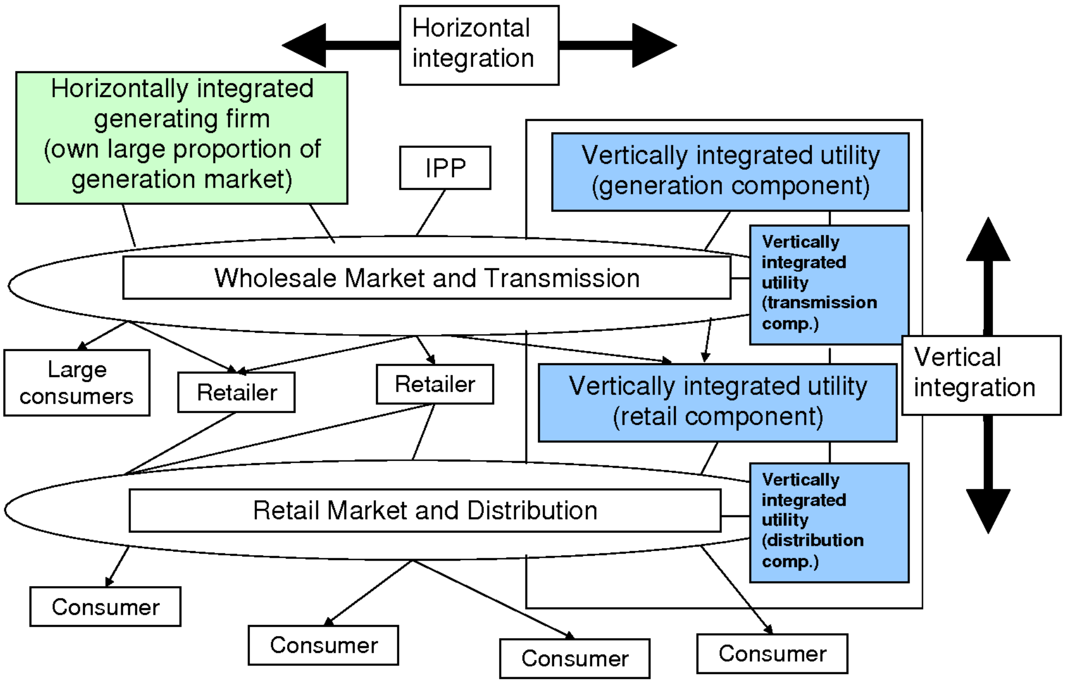


Figure 2.1: *The wholesale market structure with retail competition. Also demonstrated are vertical and horizontal integration. Based on [5].*

erated must equal demand at all times. Furthermore, the physical properties of electricity mean that system instability can arise in a matter of seconds and thus markets must facilitate (and ultimately make way for) the proper operation of the power system. Consequently, electricity must be traded in advance of physical delivery. These trades can be divided into categories based on the lead time from the price agreement to product delivery, market operation and product type. More precisely, 'energy-only' markets consist of prompt delivery markets, power exchanges, and in some instances bilateral trading between generators and retailers (Fig. 2.2). These mechanisms are described below.

2.2.1 Bilateral trading

Bilateral trades between participants is a standard feature of most commodity markets, and is also present in many electricity markets. These trades are characterised by the duration and quantity stipulated in the agreement. Forwards and futures contracts (or structured contracts) allow electricity trading over longer periods (a year or more) sometimes without formal price disclosure and can be customised to allow for flexibility. Because of their high transaction costs,

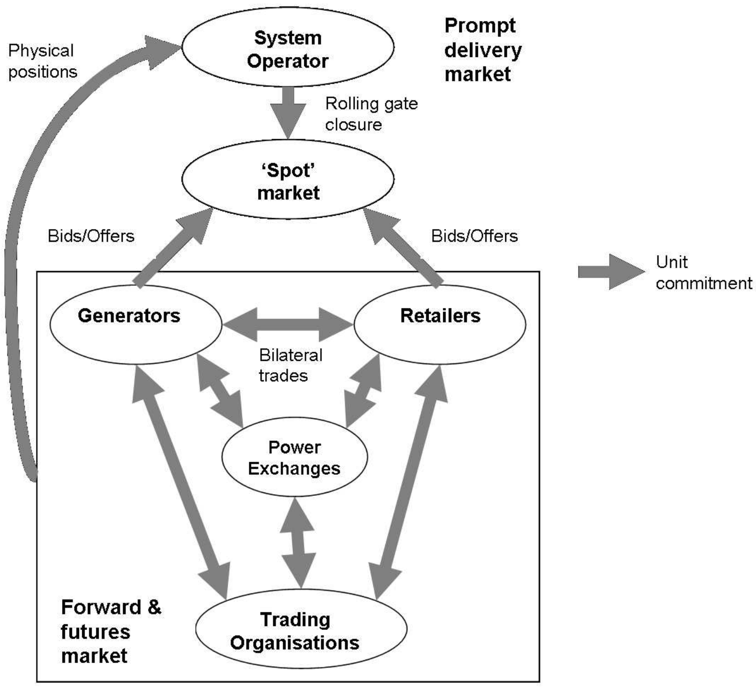


Figure 2.2: *Overview of financial and physical flows under an 'energy-only' market. Layout inspired by [7].*

long-term contracts are only attractive to those who wish to trade large volumes of energy [5]; hence they are attractive in highly vertically integrated markets.

Forward trading occurs *over-the-counter* (OTC); these trades are custom-made and can involve shorter durations and less volume, and prices are not disclosed. These take place on a variety of different platforms and as a result price discovery is reliant on a range of sources including price reporters and informal market intelligence [8].

Futures are a more transparent form of trading and take place on *exchanges*, where formal bids and offers are submitted by buyers and sellers with volumes bought and sold at a clearing price. Participants do not know the identity of the bidder or offerer, the price is there for all to see. This form of trading via auction usually goes up to the day-ahead stage and it is the job of centrally managed markets to address the prompt delivery phase.

2.2.2 ‘Spot’ markets

‘Spot’ markets for electricity are run by the SO who combines its knowledge of the network, accepted bids and offers, forecast demand, generation availability and security constraints to determine any adjustments that must be made. Bids/offers are submitted to decrease/increase production or increase/decrease consumption. These are accepted or rejected accordingly, and system balance is achieved. These markets tend to start at the day-ahead stage and work inward toward ‘real-time’ (meaning half-hour ahead). They allow participants to adjust their volumes closer to real-time (because of changes in demand forecast or unforeseen generator contingencies) and also ensure that the agreed trades are physically feasible in respect to the electricity network power flows.

2.2.3 Ancillary markets

All power markets require ancillary services. These additional services are necessary to ensure the efficient production of reliable, high-quality power [3] and that supply and demand match at all times. In most cases, there are separate markets for these services, although some parts of the service (e.g., frequency response) are mandated as part of Grid Codes. The main types of ancillary service are listed below; a more detailed description can be found in Chapter 3-4 of [3]:

- *Frequency response*: This service is concerned with the real-power balancing of the system.

- *Response and reserve*: This service is made up of fast response (spinning reserve) which can respond (i.e., increase generation or reduce consumption) within a matter of seconds. Reserve services are also included here; these provide additional back-up which can be made available in tens of minutes. Both services come together to maintain system security and real-power balancing.

- *Reactive power*: Although currently only a small service within most ancillary markets, it is vital for voltage stability. This service may become more prominent with the increasing connection of new types of generator configuration.

- *System security*: A broad area of the ancillary market, it includes services to maintain system security in both pre- and post-fault conditions. Generator curtailment agreements, generator or load intertrips and black start are key services.

2.2.4 Regulation

Most markets are imperfect, which may or mat not be corrected. Some markets require regulation to guide them toward classical 'perfection' and some, regulated or not, will never match the textbook model [9]. Power markets are an example where there is a strong case for regulation. The regulatory framework is established collaboratively by government, the regulator and the SO [1]. The regulator must be independent so that government legislation can be implemented in a transparent manner and can be easily followed or challenged. The term 'deregulated' is often used in the context of restructured competitive power markets. This term is misleading as no power market, or indeed major financial market, exists today that is not regulated in some form. The main priority of the regulator is to protect consumers, by preventing power shortages as a result of inadequate investment, or high energy prices because of market inefficiencies or manipulation. This role is of significant interest for the dynamics of capacity investment in liberalised energy markets and is discussed in detail in sub-section 2.6.5. Further, it is widely accepted that transmission and distribution operation and ownership (cf. Fig. 2.2) should remain monopolies. Therefore, in the absence of competition, the regulator must ensure that these elements also operate efficiently.

2.3 Key economic concepts

Moving on from a general discussion of liberalised electricity markets, this section introduces some fundamental economic concepts that underpin market design and provide a foundation for the work undertaken here. This review is by no means comprehensive, and the interested reader should consult Kirschen and Strbac [5] or Stoft [3] for a fuller explanation.

2.3.1 Supply and demand

In economics, the *inverse demand function*, $D^{-1}(q)$, and *inverse supply function*, $S^{-1}(q)$, illustrate the interaction between supply and demand by relating quantity, Q, to price, Π (Fig. 2.3). More precisely, the curves show the price that demand (resp. supply) is willing to pay (resp. sell) in exchange for a certain quantity. Or more intuitively, the *demand function*, $D(\pi)$, and *supply function*, $S(\pi)$, provide quantity for a given price π. Both representations are included for completeness, though when representing these interactions within a mathematical model, the standard is to use inverse functions. Whether it is better to model producers as choosing prices or quantities as strategic variables has been debated in economics for many years [10].

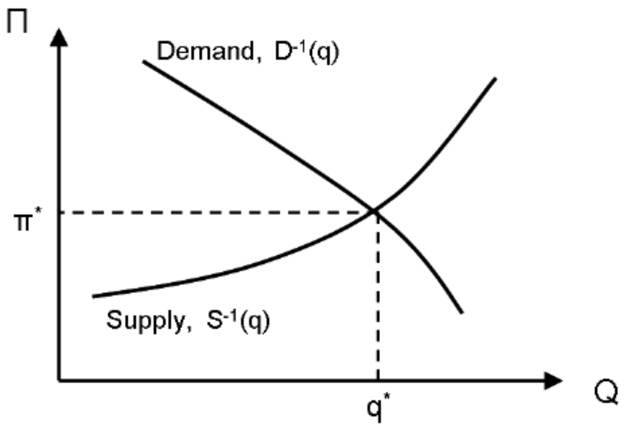


Figure 2.3: *Example of market equilibrium. Based on diagram in Chapter 2 of [5].*

2.3.2 Price elasticity

In most commodity markets, demand for a product responds to price. If the price increases, consumers will buy less and will seek cheaper alternatives or use less. The *price elasticity of demand* is defined as the ratio of a small change in quantity to a relative change in price [5]:

$$\epsilon = \frac{\frac{dq}{q}}{\frac{d\pi}{\pi}} = \frac{\pi}{q}\frac{dq}{d\pi} \tag{2.1}$$

where q is quantity and π is price. If ϵ is negative, then demand is said to be *price responsive* (the more negative, the greater the *elasticity*). Intuitively, $\epsilon < 0$ because a change in price, $d\pi$, will always result in a opposite sign change in quantity, dq. Therefore it is common to re-define demand elasticity to be positive by taking the absolute value of ϵ. ϵ is invariably positive for *price elasticity of supply* because a price change, $d\pi$, will lead to a same sign change in quantity, dq.

A characteristic of electricity markets is that the demand-side of the market is almost completely unresponsive to price. This *inelasticity* arises as demand (or consumers) neither sees nor pays the "real-time" price. For example, residential loads pay either a flat rate or one that changes between peak and off peak hours, e.g., Economy 7 in GB. Some large industrial consumers offer price responsiveness through interruptible contracts (e.g., [11]), however these remain insignificant relative to total levels of demand. The long-term elasticity of the demand side is more noticeable; if prices in the retail market increase, then consumers are likely to reduce energy consumption or switch to a cheaper provider. This change happens over a long period of time and so the short-run price elasticity of demand is usually very low (or zero). That said, with demand response and dynamic pricing likely to become significant in the future, this characteristic may change (cf. sub-section 2.7.4).

2.3.3 Short-run marginal cost in an 'energy-only' market

The marginal (or opportunity) cost is the change in total cost when the amount produced changes by one unit. It applies to both the demand and supply side of a market. The marginal cost of supply is derived from the production cost function. This could be *short-run*, where the cost of producing an additional unit is incurred by increasing output from existing facilities. Or it could be *long-run* where new facilities must be built in order to produce the additional unit. This section looks at short-run production with long-run production tackled in sub-section 2.4.3.

In power systems economics, the *short-run marginal cost* (SRMC) is incurred when producing an additional unit of energy (e.g., £/MWh). It is derived from generator *variable operating costs* using input-output curves. If the variable operating costs can be represented by a differentiable function, say $C(P)$, where P is power produced, then its SRMC is given by $\frac{dC}{dP}$. Typically, generator variable operating cost functions are convex because as production increases, the amount of input fuel required per additional unit of output power increases (law of diminishing returns). As a result, the generator SRMC function is monotonically increasing with production [5]. An example variable operating cost (a) and equivalent SRMC (b) functions are shown in

Fig 2.4. Also plotted is a piecewise linear approximation for both curves, which is a reasonable approach when the variable operating cost function is estimated from measurement data of various levels of output, and where a smooth curve cannot always be produced [5]. Also shown on the SRMC function (right) is a typical characteristic of power markets, whereby once supply hits the capacity limit, the supply curve takes an infinite jump upwards. This prevents the generator from being dispatched above its rated capacity because it is uneconomical or physically impossible to do so (discussed in detail in sub-section 2.3.6).

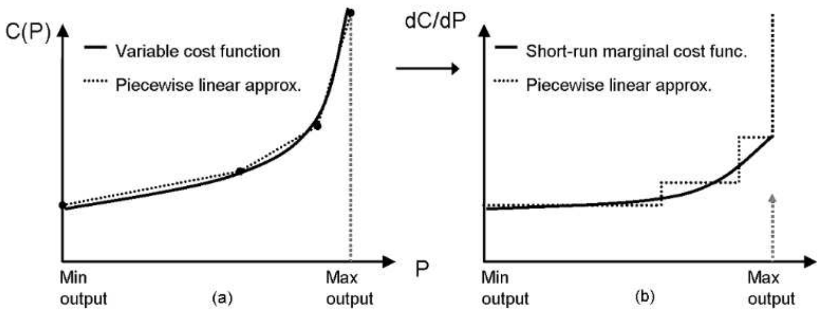


Figure 2.4: *Example of (a) variable operating cost function, and (b) equivalent short-run marginal cost function. The linear approximations are also shown.*

2.3.3.1 System merit order

A typical generation mix will contain a variety of variable operating costs and ordering individual generators' by increasing SRMC provides the system *merit order*. Summing these horizontally, provides the *aggregate supply curve* [3].

The system merit order can be broken down into three main categories:

- *Base-load* generators are expected to operate during most hours due to their low variable operating costs.
- *Mid-merit* generators are not expected to run in all hours and provide operational flexibility.
- *Peaking* generators are called upon for a small number of hours per year typically during high demand hours. They have high variable operational costs due to low thermal efficiency and running on expensive distillate fuels [12].

The duration of operation for each category is discussed in sub-section 2.4.3, but a brief demonstration of how these characteristics define the system the merit order is provided. Consider the simple example is shown in Fig. 2.5(b); each flat segment of the supply curve represents a different SRMC (indicated by $SRMC_{base}$, $SRMC_{mid-merit}$ and $SRMC_{peaking}$ on y-axis). Here linear approximations of variable cost functions have been used for all generators leading to a stepped aggregate supply function.

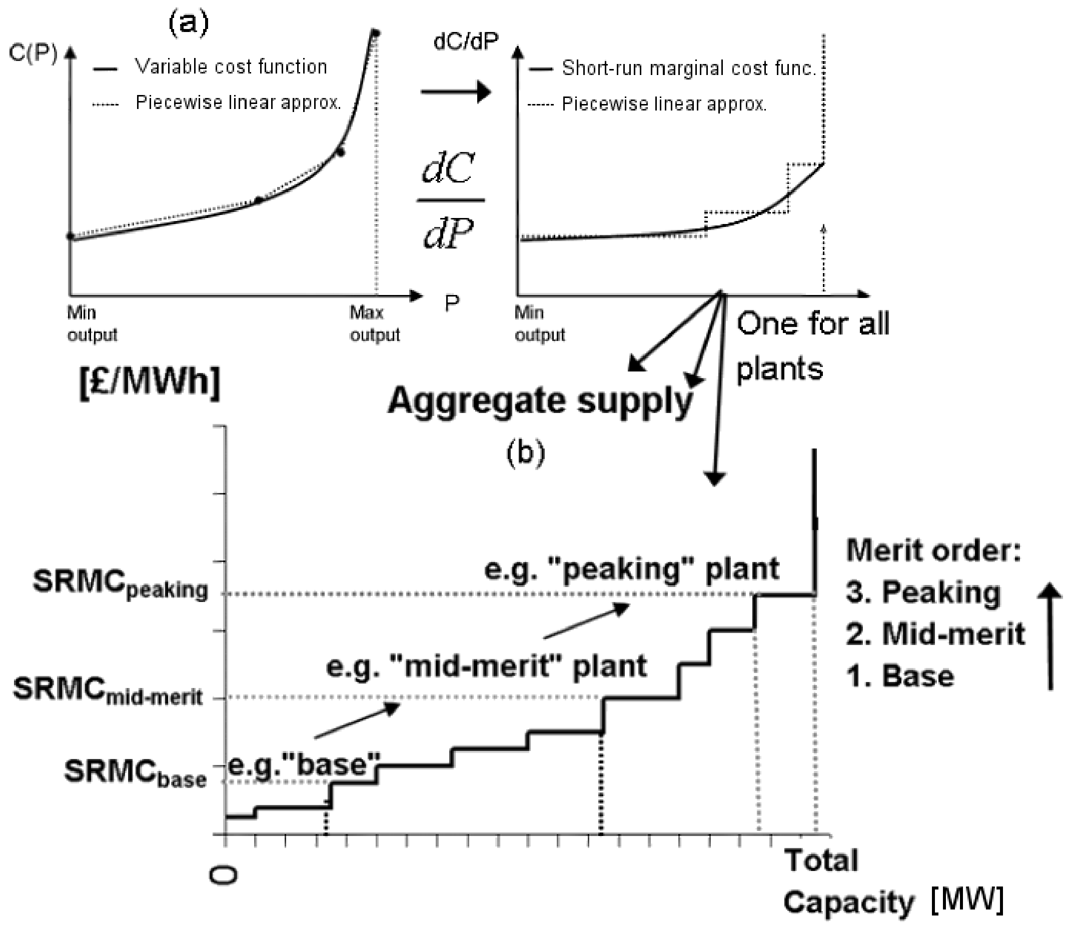


Figure 2.5: *Example of linear approximation of (a) variable operating cost function, which is used to derive (b) the aggregate supply curve. Examples of base, mid-merit and peaking capacity also shown.*

2.3.3.2 Value-of-lost-load

But what happens when demand exceeds available supply? In this example (Fig. 2.5) the value of the supply curve is infinite for quantities (MW) above total capacity. In order for the supply curve to remain finite for levels of demand above available capacity, the marginal cost of the demand side must be considered.

The SRMC of demand determines the change in total cost from consuming one unit more or less of a product. This change in total cost is typically positive if consuming more and negative if consuming less. If negative, the SRMC of demand can be thought of as the amount of compensation required by consumers to reduce their demand. If the market cannot be cleared because the available generating capacity is not adequate to satisfy demand, and to avoid load shedding (or worse, blackouts), the price must rise to a level necessary to ration demand to available supply. This price can be defined *"as the marginal cost of power at that time, set not by variable operating costs but by the opportunity cost of a consumer that has decided to reduce its demand"* [12], namely the *value-of-lost-load* (VOLL). The inelasticity of demand (2.1) means VOLL is typically very high and there is a willingness of retailers (on behalf of consumers) to pay up to the VOLL when there is a shortage in order to avoid disconnection without notice. Fig. 2.6 shows how the system supply function is adapted to include VOLL; there is now a limit to the (previously infinite) final vertical segment at maximum capacity.

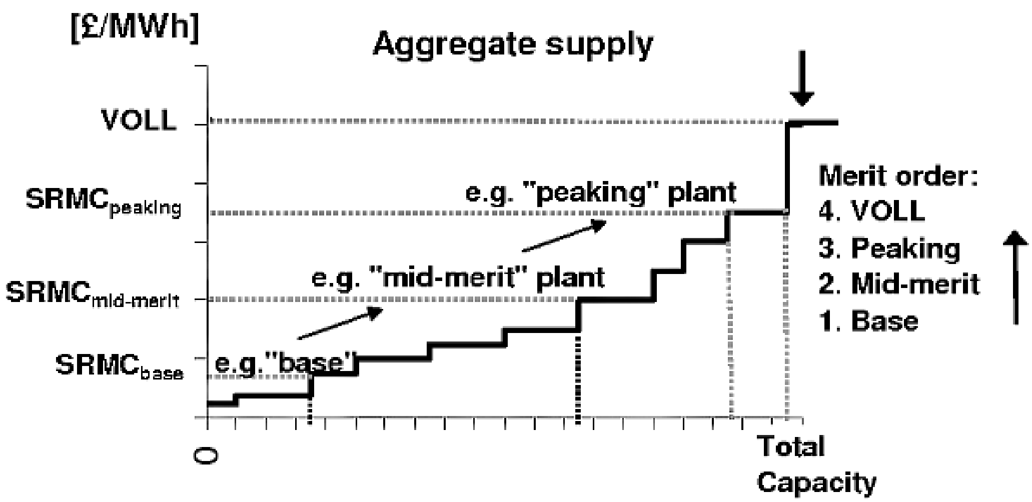


Figure 2.6: *Example aggregate supply function with VOLL.*

The issue with the VOLL discussed at great length by market designers, is what is the VOLL exactly? Depending on its use, electricity is valued very differently. For example, a hospital will place a much higher value on its energy supply than a domestic user running a washing machine. Furthermore, the load centers are highly integrated, with currently no method of disaggregation within a region. Therefore it is largely impossible to provide different levels of reliability within a geographical region and defining a single VOLL figure remains a contentious issue.

2.3.4 Perfect competition

The point where demand and supply intersect is known as the *market equilibrium* (see Fig. 2.3). If suppliers are price-takers, i.e., they cannot influence the market price by performing strategic actions to increase their market share and/or revenue, then the market is *perfectly competitive* and the equilibrium price (π^* in Fig. 2.3) is the *competitive equilibrium price*. In a perfectly competitive electricity market, the market price, π, is set by the SRMC of the last generator to be dispatched in the *economic dispatch*, $\pi = SRMC$. This is referred to as the *system marginal price* (SMP), or alternatively, the *system marginal cost* (SMC). The economic dispatch is derived from the system merit order and provides the SO with the optimal unit commitment regime (cf. Section 2.2). Because it must account for network constraints and generator ramping restrictions, unit commitment dispatches plant as economically and *efficiently* as possible and may not follow the system merit order precisely.

Not all forms of generation are dispatchable. Dispatchable generation has full operational control over its input-output. Non-dispatchable (or *must-run*) generation has limited operational flexibility, due to component preservation (e.g., nuclear) or fuel availability (e.g., wind). Consequently, commitment schedules typically start with non-dispatchable generation and work up though the merit order. This does not cause an inefficiency in the wholesale market because their SRMC is usually very low or zero and so are typically considered base-load. However an *externality* arises as a result of non-dispatchability when either i) production needs to be *constrained-off* as a result of, for example, insufficient demand or network constraints or; ii) production exceeds expectations as a result of, e.g., higher than forecast wind production. In case i) the generator will be paid its offer price, which is typically very high, and this is the *negative externality*. In case ii) the generator will be providing cheap (and sometimes 'free') electricity and thus displaces a generator with a higher SRMC. This benefit of a reduction in total production cost is the *positive externality*. Note that in a well-functioning market, when constraining-off dispatchable generation, the bid price should be greater than the offer price (e.g., in the case of fossil-fuel generation, from savings made on fuel not burnt), thus reducing total production cost. This is not always the case, but a detailed discussion is beyond the scope of this thesis.

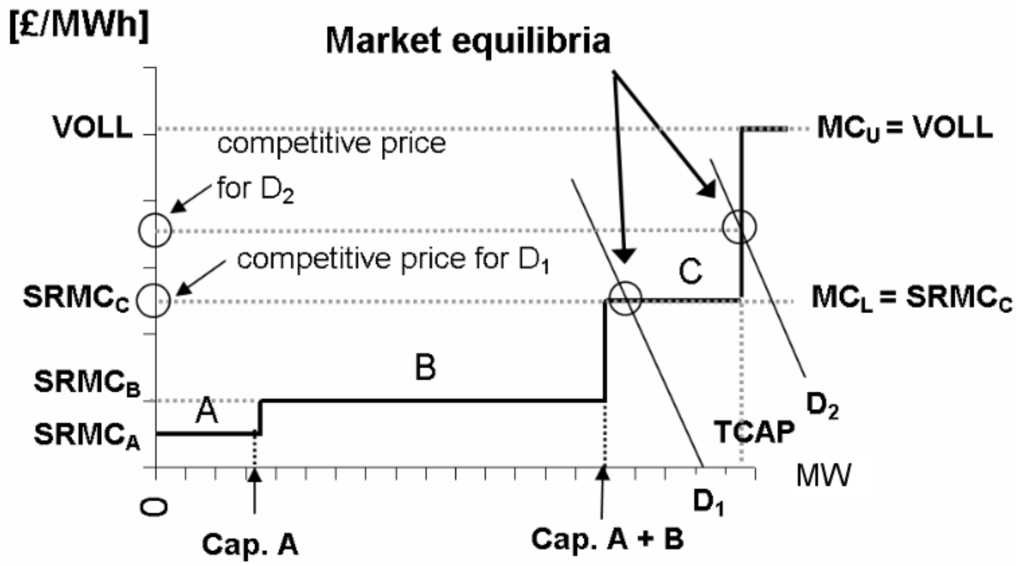


Figure 2.7: *Example aggregate supply function with competitive market equilibrium for two levels of demand.*

Two example demand curves and equivalent competitive equilibria are shown in Fig. 2.7. The leftmost curve, demand D_1, intersects a flat portion of the supply curve and the rightmost curve, demand D_2, intersects a vertical portion. The steep gradient of demand indicates low demand elasticity. A discontinuous supply curve means that the SMP can take on a *range* of values for the same level of supply (x-axis). In this event, competitors will adjust their output so that the SRMC-range contains π, i.e., $MC_L \leq \pi \leq MC_U$, where MC_L is the savings from producing one unit less of output (e.g., $SRMC_C$ in Fig. 2.7) and MC_U is the cost of producing one unit more (e.g., VOLL in Fig. 2.7) [3]. This change in outputs makes intuitive sense in a perfectly competitive market because price-taking generators will reduce their output at a cost of π, saving MC_L, as long as $\pi < MC_L$ or increase it if $MC_U < \pi$. This means that the competitive market price never exceeds marginal cost and the fundamental rules of economics hold [3].

With the perfectly competitive market model based on the SRMC of generation, there is no explicit mechanism to cover startup and no-load costs, though some markets have been designed with this in mind. For instance, under a classical pool system, *side payments* are made to generators with accepted bids when the pool price is insufficient to cover start-up and no-load costs (known as *quasi-fixed costs*) [3]. Some exchanges tackle this problem by having a multiple bid structure where generators bid in a price and a minimum run-time. There is some debate as to which design is better, yet the relative infancy of markets makes it difficult to determine which is most efficient. Furthermore, as noted in [3], the cost of committing generating units is about 1% of retail bills and although generators may change their commitments closer to real-time if the price is too low, they are more likely to do it in low demand hours when there are adequate alternative resources available.

2.3.5 Revenue and profit

For markets to endure, all parties must benefit, for example producers must receive profits and consumers must experience good value for money. In economic terms, these are known as supply and demand *surplus*. If the demand and supply function are well-defined then, given market price, π, these values can be quantified. Furthermore, if the market is perfectly competitive, then the market equilibrium maximises the producers' and consumers' surplus. The *gross* and *net surplus* for consumers' is shown if Fig. 2.8(a), with similar concepts for producers' *revenue* and *net surplus* shown in Fig. 2.8(b).

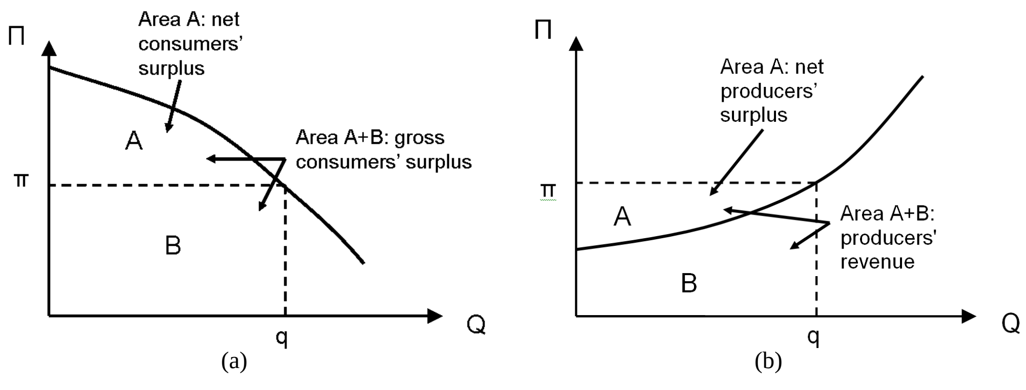


Figure 2.8: *(a) Consumers' surplus and (b) producers' surplus. Plot inspired by [5].*

Scarcity rents for producers' are an important concept in power system economics. Taking the definition used by Stoft [3], they are revenue minus variable operating cost. Equating this with the concepts described above they are the producers' (short-run) net surplus. Subtracting startup and no-load costs from scarcity rents provides *short-run profits*. Note that an alternative definition of scarcity rent exists, termed the *"folk definition"* by Stoft [3], where scarcity rent is defined as revenue minus the highest revenue earned before total generation becomes scarce (Area 3 in Fig. 2.9). For the remainder of this thesis, the first definition is used.

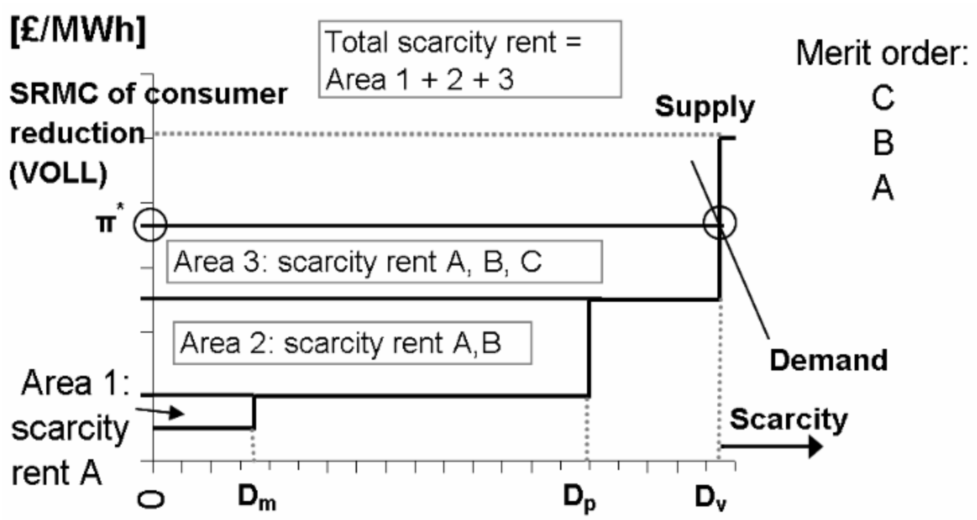


Figure 2.9: *Plot demonstrating how generators earn scarcity rents.*

Using a simple supply curve, Fig. 2.9 demonstrates how at very high levels of demand, all generators earn scarcity rents. All available generation is producing at full output and demand is willing to pay up to the VOLL for additional production. When the market clears at a price level above the SRMC of the most expensive generator in the merit order (i.e., $MC_L \leq \pi \leq MC_U$ in Fig. 2.7), this is termed a *price spike*. At other times generators gain scarcity rents in periods when the market price is above their SRMC.

While standards of generation adequacy remain high, there is a need for adequate volumes of peaking capacity to be connected to the system. If the market is perfectly competitive, peaking units will only gain scarcity rents from price spike periods. In a well-functioning market, price spikes occur during a small number of, typically high, demand periods. The uncertainty surrounding the height and duration of price spikes makes the recovery of fixed costs (FCs) uncertain and investment in peaking capacity a risky prospect for investors. Moreover, in order to recover invested capital and FCs, base-load generators require scarcity rents in the majority of operating hours. Mid-merit generators can be economic even with a lower utilisation factor, provided they gain scarcity rents in an adequate number of hours.

There is no fundamental limit to how high these price spikes can go, however some markets do impose a price cap (e.g., Australia's National Electricity Market is capped at 10,000 \$/MWh). A cap ensures that energy buyers do not continue to pay higher and higher prices for energy that is not available. This cap is the maximum amount that any generated MWh of electricity can be purchased for. As suggested in Fig. 2.9, a good benchmark for this cap is the VOLL. Price caps appear later when regulatory issues are discussed.

2.3.6 Market power

Sub-section 2.3.4 described the characteristics of a perfectly competitive market. Yet competition is not always perfect and participants may be able *"to alter profitability prices away from competitive levels"* [3] and exercise *market power*. It can be described as a three step process [3]: i) exercise; ii) effect on price and quantity; and iii) impact on market participants. The first point is an important one in terms of market regulation, because participants may *have* market power, yet they may not actually *exercise* it. That said, it is a fundamental assumption of economics that if a participant has market power, then the rational form of action is to exercise it [3].

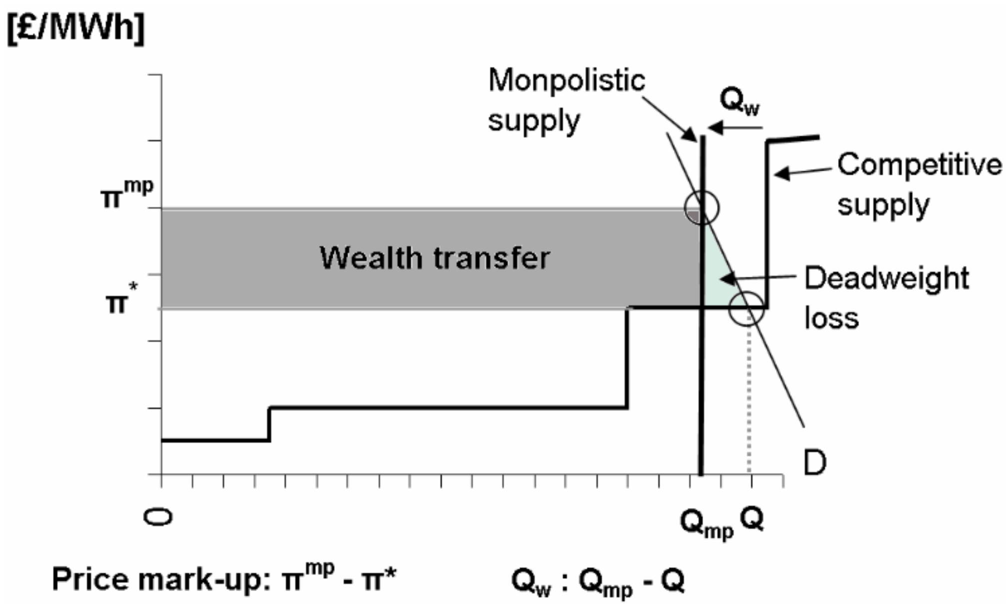


Figure 2.10: *Example of quantity withholding and price mark-up. Breakdown of wealth transfer also shown.*

Market power works in the following manner: Firstly, the characteristics of participants with the ability to exercise market power must be defined. To keep things simple, this explanation is restricted to the supply-side of the market, which is reasonable given that generator market power is of primary concern to ESI regulators and market designers. Consequently, the ability to move prices *above* the competitive level is of interest. This is termed *monopoly power*. Generators with monopoly power have a large share of the generation market and *quantity withholding* is the primary method of increasing profitability. By withholding capacity, the supply curve is shifted to the left and the market price is higher than in the perfectly competitive case. This defines the effect on quantity and price (point (ii) above). An example of this is shown in Fig. 2.10. The revenue received from the *quantity withheld*, Q_w, is zero, and the higher market price, π^{mp}, results in an increase in profitability across all remaining dispatched capacity, which as a monopoly firm, the generator will own a large share of. This increase in price $\pi^* - \pi^{mp}$ is referred to as market price *mark-up*. Submitting a supply curve above marginal cost is often equivalent [13]; both move the price away from the competitive level and hence increase profitability. A third method of exercising market power is the exploitation of network congestion to raise prices in a particular location [13]; this is not of primary concern here.

The potential for generators to submit a supply function well above marginal cost is more evident in peak demand periods due to the lack of alternatives. Generators submit a normal supply curve for most of their capacity, apart from the last few MWs which are bid in at an extremely high price (“hockey-stick” bidding). This price *mark-up* during peak demand hours occurs because firms can raise their bids knowing that the lack of alternative resources will mean bid acceptance. This has been observed in US [14] and more recently in European markets [15].

Industry regulators are concerned about market power because it results in a transfer of wealth from consumers to producers. In economic terms, this means a reduction in consumer surplus comprising of *welfare transfer*, which is the price *mark-up* times the total quantity produced, and *deadweight loss*, which is the loss as a result of market inefficiency (Fig. 2.10) [3]. This describes the final step of the process: the effect on market participants (point (iii) above). Methods to measure market power are discussed in Section 4.5.

2.4 Generation investment in market environments

So far the discussions of power markets has focused mainly on short-run issues and this thesis is concerned with electricity market dynamics over a long-term time frame. Therefore it is useful to consider the long-run dynamics of supply and demand, in particular the mix and amount of generation over time required to maintain generation adequacy standards.

2.4.1 Standard of generation adequacy

In countries with a reliable electricity supply industry, electrical energy is viewed as a public good that should always be available, and standards of system adequacy in the power industry are high. System adequacy is a broad topic, however this thesis focuses on *generation adequacy* only. This is the existence of enough generation capacity installed or under construction to meet system demand in the long-term whilst considering plant maintenance, unscheduled outages, utilisation factors, variable generation and unpredicted contingencies [16].

Expanding on this, plant maintenance is a type of *unforced outage*, i.e., a schedulable outage that can be shifted in time if required. A *forced outage* is the result of a disturbance (e.g., generator fault) which is not planned. The possibility of such an event is known as a *contingency* [3]. Given this, a capacity margin over theoretical peak demand is needed. The available methods for assessing generation adequacy are discussed in Section 2.8.

In order for adequate capacity to be maintained in the face of demand growth and retirement of existing capacity, investment in new capacity must be forthcoming. Moreover, to maintain an acceptable level of security of supply risk and keep prices within reason, this investment must be timely and efficient.

2.4.2 Competitive market-driven investment

In a perfectly competitive power market, the competitive price is established by finding the marginal resource that balances supply and demand [1]. Generators receive short-run profits (or scarcity rents) in periods when the system price is above their SRMC. In order to cover their total fixed cost of capital, these profits must be positive on average. The market is said to be in *long-run competitive equilibrium* if 1) each generator (or type of capacity) receives, on average, scarcity rents that are equal to their fixed costs, and thus total profit for all generators is zero in the long-run, and 2) the total cost of supply is minimised [3, 17]. More precisely, if scarcity rents are insufficient to cover fixed costs (long-run profit below zero), then generators will not invest. Conversely, when supply tightens or demand expands, price spikes increase scarcity rents and this is viewed as an investment signal to generators [18].

2.4.3 Demonstrating the minimum cost of supply

Firstly consider demand. The most common method of describing demand in the long-term is via a *load-duration curve* (LDC); the LDC is equivalent to the load-over-time curve sorted in order of decreasing power. This is demonstrated in Fig. 2.11, where the aggregate load time series in GB for 2010 is shown above the equivalent LDC. Note the maximum duration of 17520, representing a full year of half-hourly loads. A LDC is made up of five main features: duration, peak-load power, base-load power, total energy (area under the curve) and shape. The first four are single parameters and the latter can be described using a table of durations and powers or using a fitted function. When constructing the LDC, unpredictable factors such as weather or shifts in consumption timings tend to cancel out.

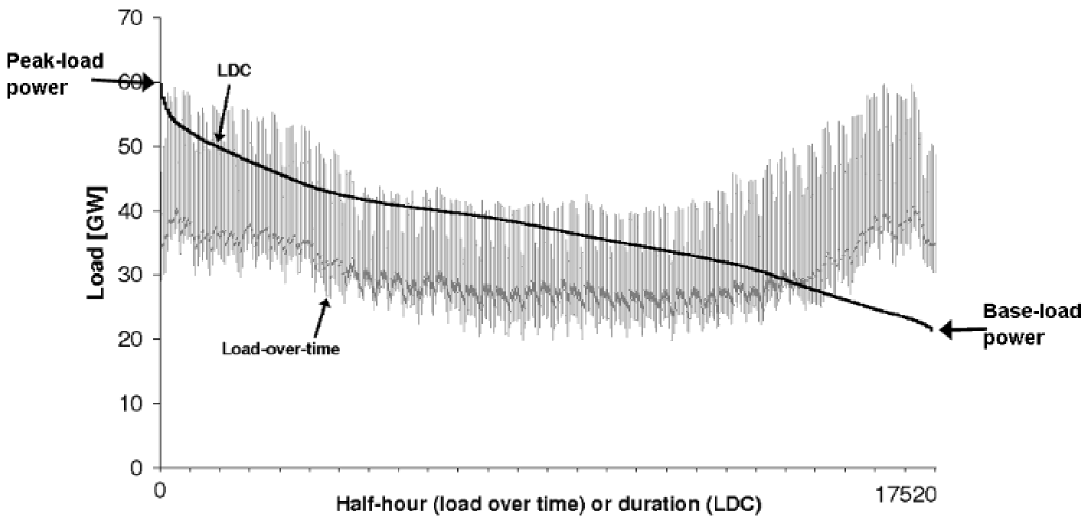


Figure 2.11: *Half-hourly load-over-time (year) for GB in 2010 (grey) and equivalent LDC (black). Source: [19].*

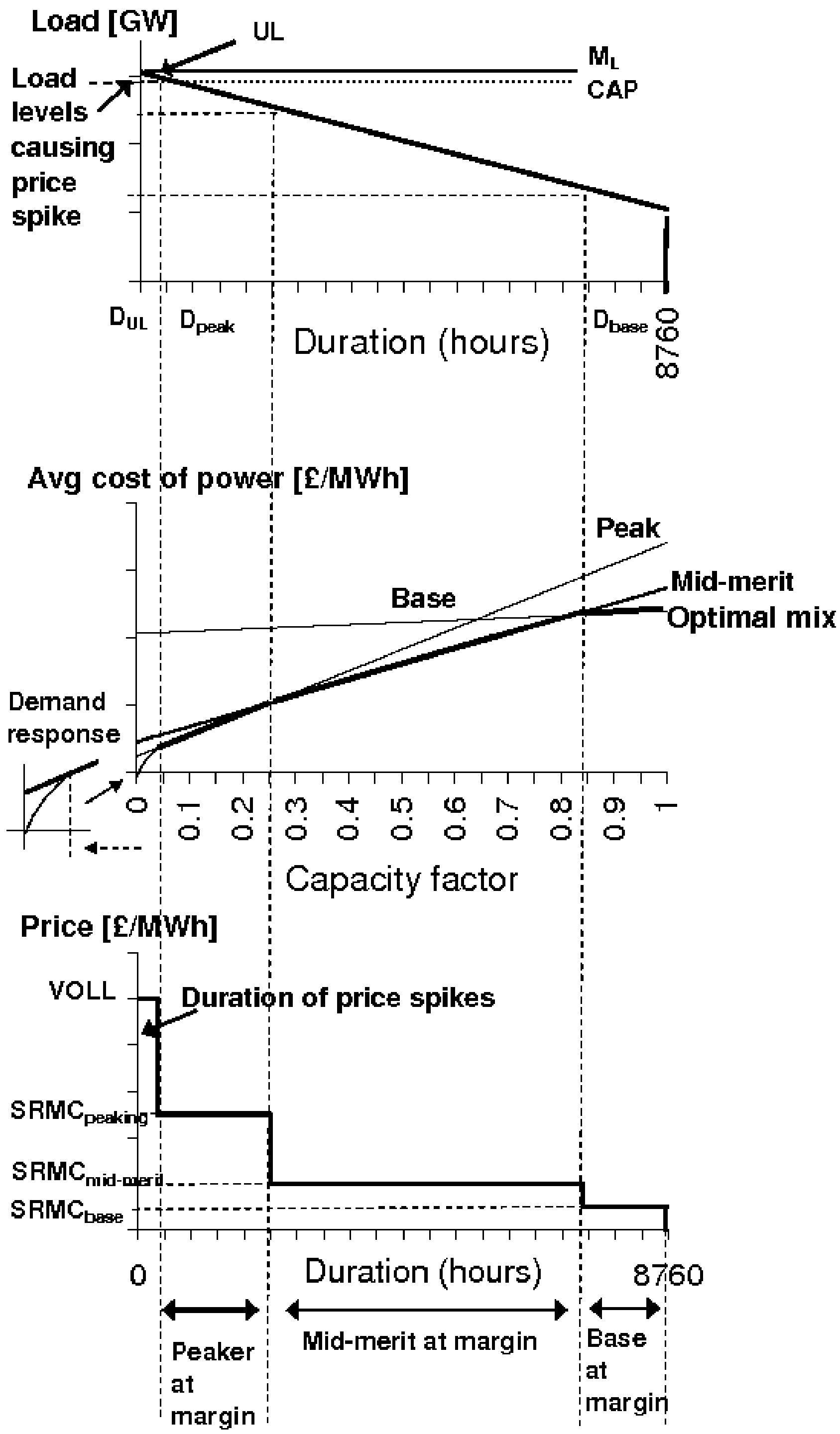


Figure 2.12: *Demonstrating the minimum cost of supply; load-duration curve (top); screening curves (middle) and price-duration curve (bottom). Optimal plant mix highlighted by solid black line.*

To demonstrate the minimum cost of supply, a simple linear LDC is shown in Fig 2.12, and the middle plot shows how the technology *screening curves* are constructed and the optimal (least cost) plant mix is determined. The solid thick black line traces the optimal mix, with sections of each screening curve used. Screening curves are a straight-forward way to compare generation costs. They account for total fixed costs (investment and operational), variable costs (VCs) and capacity factor (CF), and are a plot of average cost (e.g., £/MW-h or £/kW-y) as a function of CF (the proportion of time the plant is expected to be operating). Moreover, variable costs are the slope, fixed costs the intercept and the area under the optimal mix curve is the total supply cost. It is important to note that this is not an average cost of energy (£/MWh), but the average cost of capacity measured in £/MWh (i.e., a long-run view) [3]. The method used here to determine the screening curves is based on Stoft [3] and a fuller explanation of the information obtained in these graphs is given by Green [12].

Both components of the generation costs must be expressed in the same units - in this case £/MWh. VCs are typically expressed in cost per unit of energy produced so no extra work is required here, though the FC element requires conversion. The technique used is to amortise the FCs over the plant lifetime via the *capital recovery factor* (CRF) for a given discount rate. The CRF is the ratio of the stream of fixed payments (annuity) to the present value of the payments over a given length of time (in this case the power plant lifetime).

The formula is given by:

$$CRF = \frac{r \cdot (1+r)^n}{(1+r)^n - 1} = \frac{r}{1-u^n} \tag{2.2}$$

where r is the discount rate, n is the lifetime of the project and $u = (1+r)^{-1}$ is the discount factor. The deferred (by one year) CRF is given by:

$$DCRF = u \cdot CRF. \tag{2.3}$$

The FC element is made up of two parts. The overnight (or construction) cost (OC) (£/kW) of capacity discounted to the first year of operation:

$$OC = p_c \sum_{i=0}^{\tau-1} M_i \cdot (1+r)^{-(i-\tau)}, \tag{2.4}$$

where p_c is the cost (£/kW), M_i is the capital expenditure vector for the project with $\sum_{i=0}^{\tau-1} M_i = 1$, as it is common for capital expenditure to be spread throughout the construction period. τ

is the construction time (years). For simplicity, OC may be left as p_c, with full construction cost assumed to be incurred instantaneously. Next are the annual FCs (£/kW) of plant upkeep, which can be transformed into a stream of fixed payments assumed to be made at the beginning of each year (annuity-due):

$$FC = p_f \cdot \frac{1 - u^n}{1 - u}, \tag{2.5}$$

where p_f is the fixed operating cost (£/kW/yr). The full amortised fixed cost (FAFC) (£/kWy) discounted to the start of the first year of operation is given by either [3]:

$$\begin{aligned} FAFC &= DCRF \cdot (p_c + FC) \\ &= DCRF \cdot p_c + p_f, \end{aligned} \tag{2.6}$$

or if considering the expenditure schedule in (2.4) then:

$$FAFC = DCRF \cdot OC + p_f. \tag{2.7}$$

The full screening curve (£/MWh) is defined as $FAFC/8.76 + CF \cdot VC$.

The three screening curves plotted in Fig. 2.12 are constructed to represent the main categories of operating range. Base-load generators are typically capital-intensive with high fixed costs, which results in a high screening curve intercept, but shallow screening curve gradient on account of low SRMC. Mid-merit generators have lower capital and fixed costs and therefore a lower screening curve intercept. By comparison variable operational costs are higher, which results in a steeper gradient than base-load. Peaking generators have a low screening curve intercept due the lowest capital costs, but have steep screening curve gradient on account of high SRMCs. Screening curves are a good starting point for comparing costs of generation, although it is important to note that they do not capture every aspect considered by investors. That said, they do demonstrate the reason for having a generation market which consists of a wide range of technology types each with varied cost and operational characteristics.

Also shown in Fig. 2.12 is the cost of satisfying load throughout the year via the *price-duration curve* (PDC). Note that this is the *long-run* marginal cost (LRMC) of generation (indicated on y-axis for each category of generation). SRMC represents the incremental cost of instantaneous adjustment in production, whereas LRMC accounts for the building of new capacity and so includes FCs. The PDC also reaches the VOLL in some hours, although the duration of this period is quite short. The PDC in Fig. 2.12 also demonstrates how in the presence of some demand response the total cost of supply (area under optimal plant mix curve) is reduced.

In the example given in Fig. 2.12, if the level of total installed capacity is given by CAP, which is below the maximum value of load M_L, i.e., there are periods where available generation is not adequate to meet required load. The area UL gives the expected annual average load that is required to be shed (GW) (or perhaps, more precisely, the duration that load is required to be rationed to available supply), and the duration of this load shedding is given by D_{UL}. In a reliable power system, typical values of D_{UL} are 0.003 or 1 day in 10 years [3]. Furthermore, the marginal price for power in these periods will reach the VOLL, and if the amount of installed capacity is in long-run equilibrium, $D_{UL} \cdot VOLL$, is just high enough to cover the fixed costs of peaking generators.

The greater the installed capacity, CAP, the smaller the value of D_{UL}. This reduces the cost of unserved energy ($D_{UL} \cdot VOLL$), although it increases the cost of serving load (cost of additional capacity). The VOLL and total FC (or FAFC (2.6)) of the most expensive peaking generator can be used to determine the optimal level of reliability, or duration of load shedding D_{UL}, from the point of view of society. The condition for optimality is given by: $D_{UL} \cdot VOLL = FAFC_{peak}$. So the optimal duration of load shedding is given by

$$D_{UL} = FAFC_{peak}/VOLL. \tag{2.8}$$

Note that this does not take into account the SRMC of the generator serving the extra energy, though given that D_{UL} is typically very small, then the extra cost incurred ($D_{UL} \cdot SRMC_{peaking}$) is small in relation to the cost of the additional capacity.

Plainly, this result is based on a very simple model of reliability. There is unlikely to be a situation where total installed capacity is below theoretical maximum load. However once generation outages and load forecast errors (Section 2.8) are considered, one can see the importance of this result. What is more this enables the long-run equilibrium to be characterised with little information [3]. That said, the VOLL is notoriously difficult to determine, and, in a market environment where cost of capital differs between participants, the FAFC can also be difficult to establish. This makes design of reliability policies challenging. This will be discussed in sub-section 2.7.3.

To conclude, Fig. 2.12 is a neat way of comparing load, production costs and price duration. It also demonstrates that both energy and capacity prices are required in order to determine the minimum cost of supply. However this representation does not consider many of the important factors such as time delays, lumpiness of new build, growth in demand, market power and changing fuel costs. These topics are discussed in the following sections.

2.4.4 Effect of market power

Although the exercise of market power occurs in the short-run, there are long-run effects of market power to consider. These are of particular interest for generation investment over the long-term. The result of price *mark-up* means more profit for generators and one of two outcomes: Either the result of increased revenues bring on investment in new capacity, which in turn feeds back into the market and lowers prices. Alternatively, more sophisticated firms who account for the effect of their investment upon mark-ups for their existing fleet may deliberately not invest in order to keep prices (including mark-ups) high. In the face of demand growth and capacity retirement, large market players may invest to maintain market share. Moreover, the system becomes reliant on large market players to undertake the bulk of investment in response to the additional profits received.

2.4.5 Investment cycles in power markets

Many existing commodity markets are known to exhibit cyclical patterns, e.g., aluminum, real estate and copper [20]. More recently, attention has turned to liberalised electricity markets. Even the most mature power markets are less than 25 years old, thus limiting the scope for an investigation into the presence of cycles. A recent study by Arango [20] looked for empirical evidence of investment cycles in deregulated electricity markets by exploring a *"cycle hypothesis"*, with capacity margin as the indicator of cycles. Autocorrelation analysis was performed on 18 years of historic capacity margin data to establish the presence of cycles in the England & Wales (E&W), Chile and Nordpool systems. For the data analysed, results show that both the E&W and Chilean systems show a strong indication of cycles, with results from Nordpool considered *ambiguous*.

Generation capacity investment is prone to investment cycles owing to the long lead time, capital intensiveness and lumpiness of new build. Time delays and lumpiness of capacity play a key role in the boom and bust cycle phenomenon. Generation investment has time lags of 3-5 years for gas and coal plants and 8-10 for nuclear before becoming operational [21]. Fig. 2.13 shows this cyclical concept in its most basic form. Tightening of supply and demand results in frequent and high price spikes, which induce a wave of new (lumpy) investments. This is followed (after time delay) by a supply increase and a period of excess capacity. This leads to lower prices and capacity withdrawal, which returns the system to a period of high prices. Thus the cycle repeats.

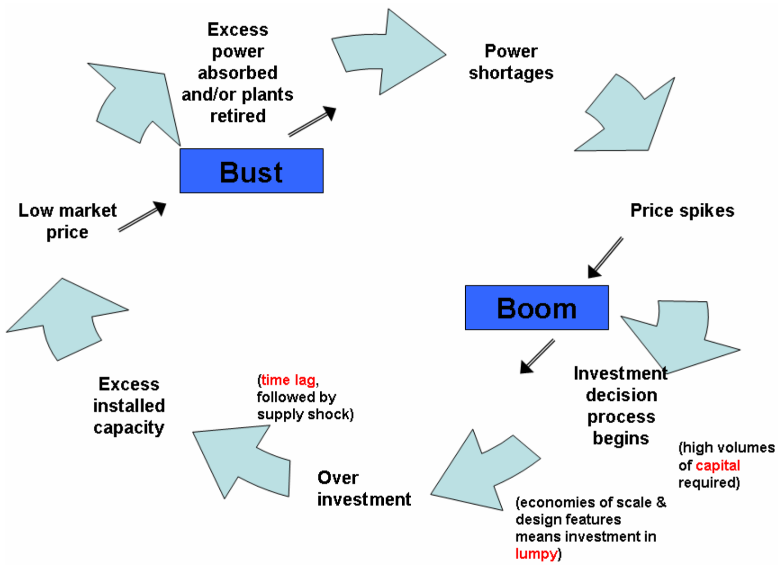


Figure 2.13: *Investment dynamics and boom-and-bust cycles in electricity generation.*

The *cobweb theorem* [22] neatly describes the interaction between supply and demand over successive periods of trade. The elasticity of total market supply and demand curves determine how quantity and price evolve as either continuous (elasticity of supply and demand equal), divergent (elasticity of supply greater than demand) or convergent (elasticity of demand greater than supply) fluctuations. Fig. 2.14 shows a convergent fluctuation; a demand of q_1 will result in price π_1. Supply will react by withdrawing some quantity to leave only q_2 available to the market. This will then result in price π_2, which is higher than π_1 on account of the supply "shortage". Equivalent examples of the patterns witnessed for divergent and continuous fluctuations are shown to the right of the main figure. The oscillations in price and quantity as a result of these fluctuations are displayed in Fig. 2.15. One can see the oscillations are damped and converge to the equilibrium after a number of time periods. The frequency of oscillation in Fig. 2.15, dt, is dependent on the supply and demand adjustment time lag.

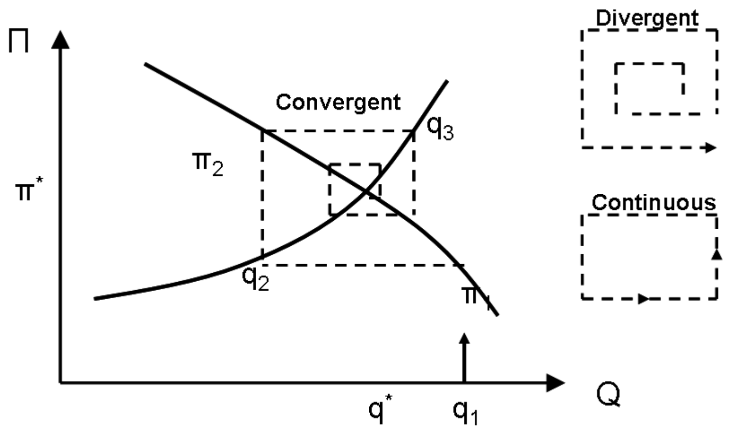


Figure 2.14: *Example of convergent fluctuation. Equivalent example pattern for divergent and continuous fluctuations displayed on right.*

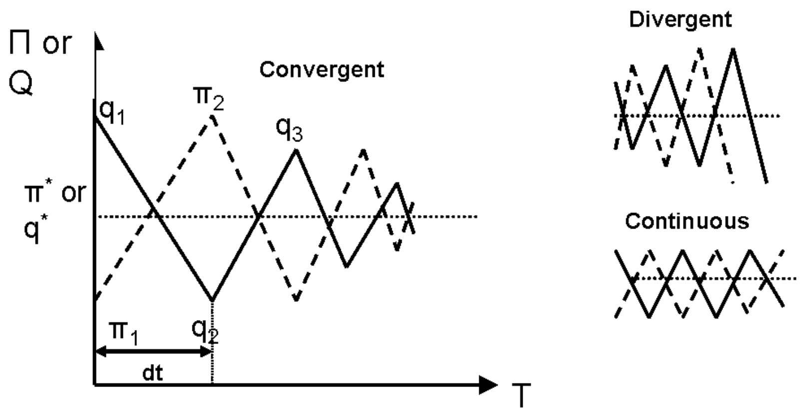


Figure 2.15: *Example of oscillations in price and quantity converging to equilibrium under convergent fluctuation. Equivalent example pattern for divergent and continuous fluctuations displayed on right.*

The cobweb theorem is used here to illustrate the characteristics of price and quantity oscillations. It is noted that in order for these dynamics to persist, producers' must form expectations about future production on the assumption that present prices will continue, and that individual production plans will not affect the market [22], i.e., producers may continue to make systematic forecast errors even if their expectations, based on historic trends, turn out to be wrong. The implications of alternative expectations hypotheses are discussed in section 4.4.1.

2.5 Costs of generation

It is worthwhile exploring the cost characteristics of electricity generating capacity in more detail. So far the discussion has been very general, with capacity discussed in terms of base-load, mid-merit and peaking generation only. In this section, some specific technologies of relevance to GB are given.

To begin, the concept of *levelised cost* (LC) must be introduced. The LC for a generating plant is the discounted cost of each unit of energy produced over the lifetime of the plant. LCs provide a straight-forward method of comparing the unit costs of generating technologies over their lifetime (e.g., [23]). They can be computed statically with a snap-shot cost analysis where costs and capacity factor are assumed constant over the plant lifetime or as a dynamic analysis where energy served and costs factors vary from year to year. The former is the more popular approach, with [24] showing levelised cost comparisons for regions across the world.

A recent GB-specific study [21], has estimated the LC for the primary technologies used in power generation (and estimates future costs and less mature technologies). Fig. 2.16 shows a breakdown of LC into its components parts for five of the technologies studied. This demonstrates how for some technologies costs are dominated by fuel costs (e.g., fossil-fuel generation) and others by capital costs (e.g., nuclear and wind). These are important considerations for any investor and the implications of these characteristics are explored in the following sections.

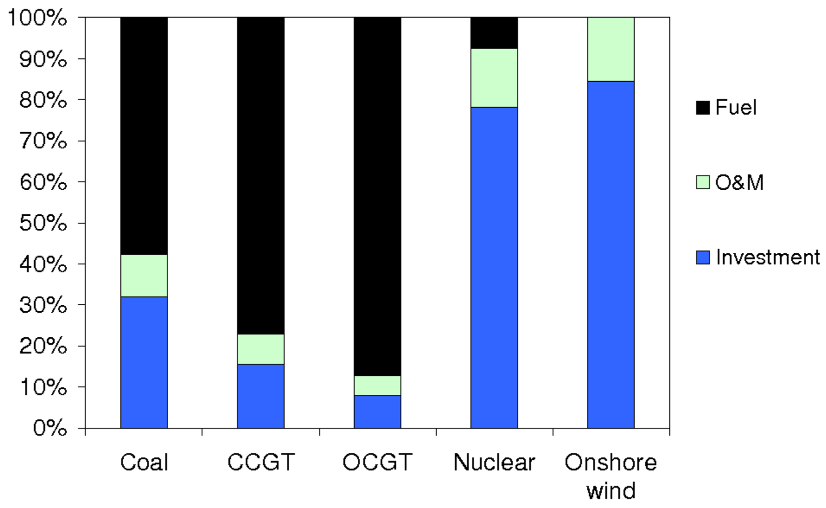


Figure 2.16: *MottMacDonald composition of levelised costs of generation for 2009 project start with 10% discount rate [21].*

2.5.1 Fossil-fuels

Coal- and gas-fired generation are the main conventional technologies used in large power plants. During the early days of large-scale power generation investment, an abundance of cheap indigenous coal (or more recently natural gas) meant that fossil fuel-based generation was relatively cheap. Although both coal- and gas-fired generation have been a stalwart of electricity generation for many years, the drive to decarbonise the electricity sector has changed their economic attractiveness. For example, the European Union (EU) introduced the Large Combustion Plant Directive (LCPD) to control emissions of sulphur dioxide (SO_2), nitrogen oxide (NO_x) and particles from large combustion plants with a thermal output greater than 50 MW. As a result, existing generators must either close by 31 December 2015 or earlier if the running limit of 20,000 operational hours is reached (which commenced on 1 January 2008) or retro-fit Flue-gas desulphurisation (FGD) equipment. New builds are required to conform with the directive. Furthermore, the development of the EU Emissions Trading Scheme (ETS), which went live in January 2005, has led to an increase in variable operating costs for these forms of generation. Expected improvements in thermal efficiencies from around 35% today to over 50% for coal and from 50% today to over 60% for combined-cycle gas turbines (CCGTs),[3] mean that some of the additional costs incurred from rising fuel and emissions prices can be partly mitigated [26]. Furthermore if these technological improvements are complimented by carbon capture and storage (CCS) on a large-scale, decarbonisation of the electricity sector is still achievable with fossil fuels albeit with continuing resource depletion issues.

According to [21], the most economic new coal-fired generators are based around the advanced supercritical (ASC) design, with the more expensive integrated gasification combined cycle (IGCC) also available. The cost premium for IGCC reflects the still largely demonstration status of this technology. However once CCS is integrated, its costs are expected to move in line with ASC with CCS [21]. In the case of gas-fired generation, CCGT designs span the full operating range (peaking to base-load). Furthermore, CCGT designs combining power and heat production (CHP) are becoming popular, with GB experiencing an increase from 0.8 GW in 2000 to nearly 2 GW in 2010 [27]. Open cycle gas turbines (OCGTs) are also available for

[3]For instance, Mitsubishis J Class gas turbine is on course to achieve 60% efficiency in combined-cycle mode [25].

use as peaking units. Other examples include diesel generators; these are the most efficient option if only a small amount of energy is required [12] and are common source of back-up generation on small island systems with poor interconnection (e.g., Scottish Western Isles).

2.5.2 Nuclear

Nuclear power plants are considered base-load owing to high capital costs and limited operational flexibility. They have very low variable operating costs, and are deemed 'must run' in order to recover capital expenditure. From an emissions perspective, nuclear power is carbon-free at the point of generation. However a recent study [28] estimated emissions to be between 7 and 22g/kWh over the lifecycle of a plant[4] compared to 380g/kWh and 830g/kWh for gas and coal-fired generation, respectively.

Nuclear power requires significant waste management during and after the lifetime of the plant. Depending on the fuel cycle, fuel replacement must be carried out every 12-18 months [28]. The radioactive spent fuel must be stored and then eventually placed in a geological repository where it will remain for thousands of years. This process adds a significant amount to decommissioning costs, e.g., the World Nuclear Association (WNA) estimates these to be around 9-15% of capital cost [29]. In contrast the GB Magnox nuclear power stations currently undergoing decommisionning are estimated to be substantially higher with a combined cost of £12 billion across 11 sites[5] totalling around 5 GW, i.e., 2,440 £/kW [30].

2.5.3 Renewables

Energy generated from renewable sources is an increasing trend in power systems. Many countries have signed up to binding targets to reduce GHG emissions and the electricity sector is expected to make a significant contribution to these goals.

There are many forms of renewable generation, some more established than others. For example, run-of-river (RRH), dam and pumped storage (PS) hydro power have been used since the

[4]This figure can be much higher if the ore quality is low.

[5]Berkeley, Bradwell, Calder Hall, Chapelcross, Dungeness A, Hinkley A, Hunterston, Oldbury, Sizewell A, Trawsfynydd and Wylfa.

1880s. PS is perhaps the closest power systems come to having storable energy, while RRH is more reliant on the day-to-day availability of the resource as there are no storage facilities. Large hydroelectric dams are popular in many countries and although expensive, they can last for many years.

Wind power generation has experienced considerable growth in recent years. In 2009, global wind energy production was 347TWh, which is about the annual electricity consumption of GB [31]. This figure is expected to increase 15-fold by 2030 [31]. In GB, the industry is developing rapidly onshore, with offshore wind expected to accelerate in the coming decade as the deadline for emissions targets approaches. Wind generation, like other renewable sources, is not only attractive because of its zero emissions but also because it is a “free” form of energy with virtually zero variable operating costs. It is non-dispatchable on account of its reliance on the availability of the fuel resource. As a result, wind is considered base-load generation. Wind capacity experiences high upfront costs with a recent report showing that onshore wind is around twice as expensive per installed kW than a CCGT plant, and offshore over 3 times [21]. These are expected to reduce significantly as the technology matures. For instance, costs of £157-186/MWh for offshore wind in 2010 are projected to reduce to £110-125/MWh for projects commissioned from 2020 (both 2010 real terms) [21].

Other forms of renewable generation used in the electricity sector include biomass, solar photovoltaics, geothermal and wave and tidal stream energy sources. The intensity of deployment, degree of maturity and level of cost varies greatly across these technologies and is resource and policy dependent. A detailed discussion is beyond the scope of this thesis.

2.6 Investment uncertainties

In competitive markets, in order to avoid the risk of losing market share and even financial ruin, firms are incentivised to improve efficiency and avoid bad decisions. Furthermore, as documented extensively in the literature (e.g., see Chapter 1 of [32]), when faced with a largely irreversible investment decision under uncertainty, whenever possible, investors will delay until more information becomes available and there is less uncertainty. These delays can result in a potential generation capacity shortfall and exacerbate the boom and bust phenomenon discussed earlier. Table 2.1 shows the key uncertainties faced by investors and the impact of these uncertainties are explored within this section.

Technical	**Market**	**Financial**	**Regulatory**	**Policy**
Construction costs	Fuel costs	Weighted Average Cost of Capital	Market design	Emissions limits
Operational costs	Demand growth	Credit risk	Regulation of transmission	Subsidies
Lead time	Emissions costs	Contractual risk	Regulation of competition	Energy efficiency
Decommissioning	Wholesale prices		Planning process	
Availability	Degree of competition			
Capacity factor				

Table 2.1: *Key uncertainties affecting firms investment decision. Source: [33] and [24].*

2.6.1 Technical uncertainty

Capital-intensive and long lead time plant are less economically attractive for investors. They will prefer technologies with short lead times that can be brought online in small incremental steps [1]. For instance, capital costs have a high impact on nuclear but less so for CCGT, conversely primary fuel cost has a high impact on CCGT; but less so for nuclear. The timing and size of other projects in the pipeline and whether more efficient technologies are available is also important. Competing firms choose different generating technologies and hence face the financial consequences of unwise investments.

2.6.2 Market uncertainty

Many investment uncertainties are introduced by competition. For example, under a monopolistic market, change is technology driven with uncertainties in demand and costs being key [34]. However monopolies are vulnerable to inefficiencies that can lead to customers overpaying for the service. In a competitive market, the introduction of multiple players means firms can no longer take prices and customer base for granted [34]. Furthermore the nature and degree of these risks will differ between project, technology and investor type. Before a build decision is taken, each of the factors must be exhaustively explored.

To address the uncertainty of fuel prices, companies may agree upstream contracts to guarantee fuel supply at a fixed cost. As stated in [23], a generator's exposure to price uncertainty depends on the technology. If a producer is a *price maker* - for example CCGTs tend to set the wholesale price in GB owing to their position in the merit order - they are largely able to pass fuel cost variations onto consumers. Conversely, *price takers* are positioned lower in the merit order and have little price setting influence.

Electricity demand in most power markets is expected to increase year-on-year (e.g., until recently, Scotland experienced a 1% annual increase [35]). That said, owing to the recent economic downturn and improvements in energy efficiency, this trend may no longer persist. Regardless of whether it is an upward or downward movement, it is very difficult to predict future demand accurately. However *"accurate long-term estimates of both power and energy requirements are crucial to effective power system planning and operation"* [35], thus highlighting the need for good demand forecasting tools.

2.6.3 Financial uncertainty

Profitability is an important factor in any investment decision. By exchanging a large sum of money today for an income stream in the future, investors will require a good rate of return. Investment risk is further exacerbated by the fact that the cost of capital is higher for private firms than for governments. Therefore investors will intuitively require a higher rate of return [36], which further exacerbates the problems described above.

Under a liberalised market framework, investment in new generation is financed by private capital. This may be done with project finance where investors create a new company for the purpose; this limits their risk because it isolates the project from existing operations, meaning lenders cannot claim against existing assets if the firm defaults on a loan. The alternative is corporate finance when new investments are financed from the company balance sheet [37]. Many forms of generation, such as nuclear and coal, are highly capital intensive in the construction phase. As a result, only large utility firms have the financial strength to take on such projects [24]. Gas-fired generation is less capital intensive and so has fewer barriers to entry; a good example being the 'dash for gas' in GB during the 1990s (and described in section 3.2.1.2). In the case of renewables, although there are no fuel or emissions costs during operation, capital costs, which account for the majority of total costs, are usually high. Perhaps a good example of this is large-scale PS; once built the system cannot do without it, yet financing it is almost never carried out privately.

In addition to technology cost characteristics, the perceived level of risk involved will differ and must be compared to alternatives (including the option of doing nothing). This concept can be explored through the weighted average cost of capital (WACC). The formula is given by [36]:

$$r = (\chi) \cdot \gamma + (1 - \chi) \cdot \kappa \tag{2.9}$$

where χ, γ and κ are the gearing ratio (debt/equity), expected bond return, and required investor equity return, respectively. The higher the risks, the higher the costs of debt and equity, and the higher the required return on investment [24]. The WACC can be used in a levelised cost analysis or net present value (NPV) analysis when revenues as well as costs are considered. When using the WACC formula to calculate the discounted costs it is important to determine whether the *nominal* or *real* rate should be applied. The formula for calculating the real rate of return is given by [36]:

$$1 + r^n = (1 + r^r)(1 + r^i) \tag{2.10}$$

where r^n is the nominal rate (2.9), r^r is the real rate and r^i is the rate of inflation.

The amount of debt involved varies between projects and size of firm. Owing to the capital-intensiveness, long economic life and uncertain revenue streams, IPPs may seek non-recourse project finance with 60 to 75 percent debt typically witnessed [38]. The debt is guaranteed by the project assets, rather than from the general assets or credit rating of the IPP. Further, [39] reports figures of 50:50 (debt:equity) for nuclear and 40:60 for both coal and CCGT.[6] This is perhaps reflective of the additional capital required for a nuclear investment over other technologies. There is a similar pattern for required equity return (κ), with nuclear requiring 15% and CCGT requiring 12%. The additional risks - and hence higher required returns - together with the financial strength required for these capital intensive investments means investment in these types of technologies can be difficult to induce in a market where less risky options are available.

The ability to secure debt will depend on the company's credit rating. Creditors will typically assess a firm's creditworthiness using rating firms such as Moody's and Standard & Poor's.

[6]Note that these ratios may differ during the construction and operational stages of the project. For example an increase in debt once the plant is operational is not unusual; the firm is more likely to secure more debt once the plant is operational as many of the risks associated with the construction stage have passed.

Rating symbols Aaa to C are used to reflect the relative credit risk of a firm. For example *"obligations rated Aaa are judged to be of the highest quality, with minimal credit risk"* and *"obligations rated C are the lowest rated class ... with little prospect for recovery of principal or interest"* [40].

Some information about company financial structure can be deduced from annual reports. Table 2.2 shows the most recent available cost of capital figures for the "big six" generating firms in the GB market. These statistics are for the whole firm, all of whom are vertically integrated in the GB market and have contrasting generation portfolios with expertise in certain technologies. Furthermore, many participate in electricity markets elsewhere. Therefore these figures may not be a fair reflection of their GB generation business or investment decisions, it is informative nonetheless.

Firm	**WACC**	**Gearing**	**Long-term Moody's Credit rating**
E.ON UK	4.2%	43%	A3
Edf	-	62%	Aa3
RWE	9.0%	65%	A2
SSE	5.4%	63%	A3
Iberdrola (SP)	-	50%	A3
Centrica	9.3%	10%	A3

Table 2.2: *GB large electricity utilities financial statistics. Derived from sources [41–46].*

2.6.4 Policy uncertainty

Policy plays a key role in generation investment in power markets. Government backing for, or opposition to, a particular technology, can make or break a technology's chances of investment. There may be policy mechanisms in place to encourage investment in cleaner technologies. These are not necessarily the best option from the investor perspective, yet if the policies are effective and the market operates efficiently, investors will choose the intended technologies and the policy goals will be met. For instance, Renewable Obligation Certificates (ROCs) are a UK Government policy mechanism used to stimulate investment in renewable generation whereby each MWh generated from a renewable source is eligible for subsidy. Other objectives concerned the environment and, in the case of Europe, were addressed through the LCPD (sub-section 2.5.1) and the Integrated Prevention Control Directive (IPCD) of 2000 [23].

In the early days of liberalisation, the main policy drive was to promote competition in the generation market and lower prices for consumers. Examples of policies that are currently hoping to influence investment decisions in GB are the Renewables Obligation (RO) and EU ETS. Both are geared toward GHG emissions reductions, the former by incentivising investment in renewables and the latter by capping on emissions by handing suppliers a finite number of emissions credits. They must trade these credits through the ETS market if they require more credits or wish to offload surpluses. The latest incentive scheme, aimed at small-scale renewable developments is the feed-in-tariff (FiT) mechanism. Already implemented in Germany for many years with success, this scheme provides a payment for small-scale low carbon installations for energy produced.

2.6.5 Regulatory uncertainty

Regulatory policy has a significant bearing on generation adequacy. For instance it determines the height and duration of price spikes [3], which as discussed earlier is the primary mechanism to induce timely investment in a perfectly competitive market. A general assumption made when implementing price caps is that the market is perfectly competitive. Some argue that price caps dampen the market signals required to bring on investment and are only implemented in order to avoid politically unacceptable price spikes [47]. It is argued in [48] that almost any price-cap level can potentially induce both too little and too much investment. If markets start to produce frequently high prices, this puts regulators under pressure to intervene and change market design.

Moreover, delays and uncertainty in changes to market regulation can also cause investors to delay building of new capacity. This is precisely what regulators wish to avoid, and may in fact be why the market reform was being considered in the first place. Therefore delays and uncertainty surrounding reform must be kept to a minimum. That said, there has been no experience of a complete 25 year capital cycle in the power sector in Europe or elsewhere, so reliable regulatory mechanisms aimed at ensuring efficient investment are difficult to develop. Some of the mechanisms that have been designed to reduce uncertainty for investors and mitigate periods of generation shortfall are discussed in Section 2.7.

The most commonly cited example of a power market failure - and one which can largely be attributed to insufficient regulation - is that of California in 2000/01. California was a unique case, yet it did demonstrate the need for clear regulation on some level and for transparency from the moment the framework is implemented. This was quite a unique event as not just one, but a series of factors came together to cause numerous blackouts which left millions without power during both the summer and winter seasons. A decade of regulatory uncertainty had led to significant under investment in capacity [49]. Additional factors which confounded the State's problems included firstly, a retail rate freeze (capped at 6.7 cents per kWh [50]) which left the retail firms close to bankruptcy once unhedged generation costs started to rise, and significant demand growth in both California and neighbouring regions (the latter reducing import capabilities) [49]. These factors led to serious congestion problems and an inability to get enough power to the load centres. It was later alleged that the energy company Enron had manipulated the market in order to make a significant financial gain from the payments received for congestion alleviation.

2.7 Markets for capacity

A "laissez-faire" approach to generation capacity investment, means no security of supply standards or planning reserve margin requirements are enforced, although regulators or system operators may identify target reserve margins [51]. Advocates of this form of market design cite that the ability of generators and retailers to enter into forward contracts for energy acts as a means of ensuring long-term security of supply. By entering into long-term contracts generators are incentivised to maintain a generation capacity capable of meeting their contracted obligations, thus avoiding having to pay penalties for non-fulfillment. Further, both generators and suppliers are risk averse and wish to avoid the costs associated with volatile wholesale prices. Forward contracting is seen as a method to hedge these risks and requires no more regulation that standard commodity markets [3]. Another benefit of forward contracting is that it reduces the potential for market power when there is low demand elasticity and a concentrated (e.g., oligopoly) generation market structure [15]. By comparison in order to make generator revenues more predictable, lower investment costs and provide adequate and timely investment in generating capacity, some markets have additional mechanisms in place. Some of the mechanisms have been implemented - with varying success - and some are still purely theoretical. A selection of these mechanisms are discussed below.

2.7.1 Capacity payments

This is a price-based mechanism. Rather than producers receiving large sums of money only during a handful of periods owing to power shortages (i.e., price spikes), they receive a smaller amount on a regular basis. These payments are typically administratively-determined and seek to achieve reliable and adequate generation capacity. They aim to cover at least part of the capital cost of new generation and encourage companies to keep rarely used units rather than decommission them [5]. The implementation of this mechanism can be split into two main categories; either it is delivered as a flat-fee independent of the wholesale price (as initially implemented in Argentinian and Spanish markets) or it is added to the wholesale price. The amount added is dependent on the amount of firm (reliable) capacity provided by generators, as in Chile where probabilistic models were used to calculate each units' contribution to overall system reliability [16].

Advocates of these mechanisms argue that the investment risk of having to rely on a small number of high price hours is removed along with the consumer risk of being exposed to very high prices for energy. Note that the argument for a wholesale price cap is stronger in a market with capacity payments. Critics argue that it is difficult to fix the capacity price: too high and consumers can end up over-paying for capacity, too low and the system could fall short. Furthermore, the definition of reliability must be clear and the metric used to determine the reliability of both individual units and the aggregated system must be transparent.

Examples of use [51]: Spain, Argentina, Chile, Columbia, Peru, South Korea.

2.7.2 Capacity markets

This is a quantity-based mechanism where a target level of installed capacity is determined by the regulator, hence the name Installed CAPacity (ICAP) markets. All suppliers which provide electricity to end-users and wholesale market customers are then required to buy a share of this requirement. Thus by stipulating a target level of installed capacity and making it an obligation of the suppliers to participate in the capacity market, this mechanism balances the supply of, and demand for capacity not met by the bilateral market or using self-supply [52]. Another benefit of such mechanisms is they can be extended to include locational signals whereby the ICAP price varies in different regions of the network depending on capacity demand within a region.

The amount of installed capacity required is typically determined by identifying a target reserve margin which is calculated based on an acceptable risk level. The capacity price is determined by the amount of capacity on offer and can sometimes be quite volatile. Further, the target risk level must be economical. For example, [53] shows how setting the level too high may lead to a cost for the expected energy served by the plant required to meet the target risk level being very high (over $6,000/MWh levelised). The problem of the target reserve margin is further exacerbated in systems with high penetrations of wind, where the calculation is complicated by the dependence of output between plants within a geographical area. Therefore robust methods for determining the optimal level of reserve margin based on standard reliability metrics are more challenging to define. The length of the period over which capacity obligations are calculated needs to be decided. Retailers prefer a shorter period as it reduces amount of the obligation they have to purchase. Having a short time-step also increases the liquidity of the capacity market. By comparison producers prefer a longer time-step as they receive more capital for investment.

Assessing and rewarding generator performance is difficult because installed capacity must be higher than peak demand and unreliable generators add to the required reserve and impose additional cost on the system. This was addressed in the Eastern US ISOs capacity markets whereby the ICAP market was replaced by the unforced capacity (UCAP) market. In this market the ISO was able to discount, depending on historical plant availability, the capacity for which it was given the credit [16]. This incentivised generators to keep a good availability record to increase the credits it would receive.

As mentioned, the prices in capacity markets can sometimes be quite volatile owing to the quantity of capacity on offer. The price paid for capacity can jump from zero to the maximum payment (typically covering peaking unit FCs) if the reserve margin is moving between being above and below the target level. To address this issue, a hybrid price- and quantity-based mechanism described in [54], which went live in PJM in June 2007 has been developed [55]. It uses a sloped 'demand curve' for capacity where the payment level varies in the vicinity of the target reserve level (e.g., +/- 4% of target UCAP). This makes the pricing mechanism smoother and less volatile compared to a stepped function which is the maximum capacity price if the system is below the target reserve level and zero above it. This mechanism is discussed, with example demand curves provided, in further detail in section 8.2.4 when it is applied to the dynamic investment market model presented in this thesis.

Examples of use [51]: PJM, New York ISO, ISO-New England, Guatamala, France, Australia's South West Interconnection.

2.7.3 Reliability requirements

As discussed above, it is possible to address reliability within both the capacity payment and capacity market designs, although specific reliability-based mechanisms are available to market designers. The most commonly cited design (and one which has been implemented in the Columbian power system) is that of the so-called 'reliability option' described in [56] and [57]. This mechanism is a financial product constructed using *"a combination of a financial call option with a high strike price and an explicit penalty for nondelivery"* [56]. Similarly to a capacity market, the regulator sets a forward target amount of capacity and a strike price for the call options. The generators then submit any number of bids to the reliability auction for the options on this target volume of power. The total of all the successful option bids is equal to the target capacity. Supporters of this mechanism argue that this coordinated entry reduces the uncertainty in achieving the target level of capacity and mitigates the boom-and-bust investment cycles described earlier [58]. The companies must also calculate the premium for their bids in order to cover potential losses when the options are exercised (i.e., market spot price exceeds option strike price). The regulator must also set the penalty fee for nondelivery of power stipulated in the option contract. A summary of its implementation is given in [18]. From a demand perspective, these instruments are attractive because there is a cap on the maximum price paid for energy (the strike price of the call option). They are also attractive from a generator perspective because they enable hedging against price volatility. From a regulator perspective, these instruments are appealing because they do not interfere with normal market dynamics; the mechanism only becomes active when the system approaches scarcity. Furthermore, they incentivise generators to only bid reliable units as the penalties for nondelivery are high. One short-coming of this mechanism is that the demand-side must decide how much it is willing to pay for reliability [5], i.e., a similar problem to determining the VOLL. This is something that varies between consumers and it is currently impossible to isolate individual loads within the system. That said, the regulator can achieve the level of reliability it believes consumers are willing to pay for by setting the amount of contracts to be auctioned at the level required to achieve "optimal" reliability.

Examples of use [51]: Columbia.

2.7.4 Alternative mechanisms

The sections above detail the most common capacity mechanisms, all of them implemented in real markets, yet there are numerous other designs available, outlined below.

In [59], a design for a *capacity subscriptions* market is given. Here customers purchase capacity (through retailers) and install a load limiting device (LLD) which limits usage to capacity purchased. The price in these contracts is modelled via a demand (for capacity) curve similar to that described in sub-section 2.7.2, by comparison the curve is defined in terms of willingness to accept demand reductions. Whenever the demand for electricity exceeds the stipulated load limit, the LLD limits consumption.

A similar approach is described in [11] where two types of product are described, both of which are active on the demand-side. The first is a *pay-in-advance* interruptible contract where retailers can agree with industrial customers to interrupt a given percentage of their load a fixed number of times throughout the contract in exchange for a tariff discount. The next is *pay-as-you-go* interruptible contract where the retailer can interrupt part of a customers load a fixed number of times in exchange for compensation. This method is not tested in terms of generation adequacy as it applies to the retail-side, if the contracts are activated during scarcity periods, this type of contracting provides the system with short-term demand elasticity and therefore partially addresses one of the main flaws of modern electricity markets.

Another option discussed in [56] is a so-called *mothball reserve* which consists of a number of mothballed old plant (typically peaking units) which can be returned to service if necessary.

In principle this method works well and can provide sufficient capacity in a relatively short-time frame compared with a new build (typically, plant can be returned to service in 6 months, compared to around 2 years for a new OCGT build). The down-side is that the longer plant remains idle, the more likely that they will be used for spares and will take longer to return to service [2]. Note that GB effectively provides this capability without an explicit mechanism in place.

Finally, by having short-term *demand response* (Fig 2.17) when there are periods of supply shortages, the SO can ration demand to available supply at a lower cost. That is, demand responding to real-time wholesale energy prices and showing significant elasticity. Furthermore the market will be more efficient and the risk of blackouts as a result of inadequate available generation and subsequent system instabilities can be reduced. For instance, the system will no longer be at the mercy of "hockey-stick" bidders when there is a generation resource shortage.

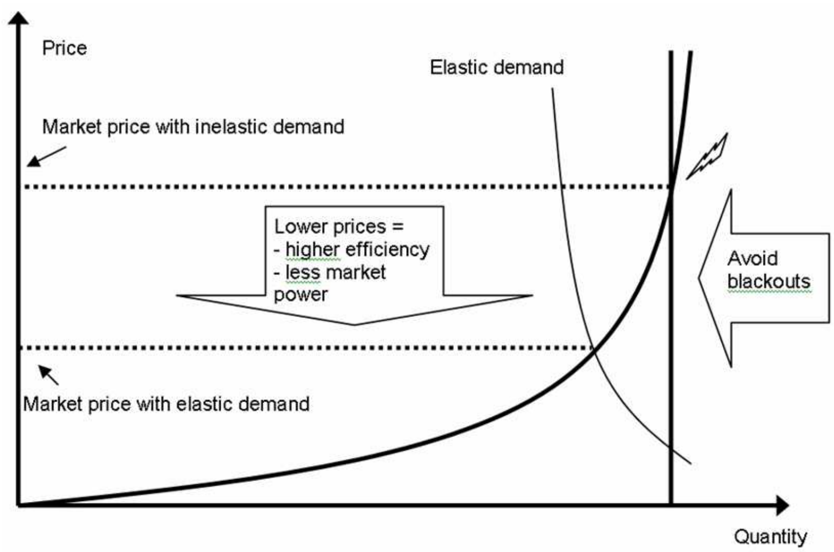


Figure 2.17: *Influence of demand response on market clearing. Based on [60].*

It seems increasingly likely that many countries will turn to demand-side participation as part of so-called "smart grids" in the coming years to increase market efficiency, reduce aggregate energy requirements and compliment increasing penetrations of wind generation. It should be noted that the simulation model presented here does not consider the effect of demand-side participation on generation investment dynamics. Possible extensions to the modelling methodology in order to represent this additional characteristic are discussed in Section 9.1.5.

2.8 Measuring generation adequacy risk

Moving on from the basis of generation adequacy and methods to enhance it, this section presents the mathematics used to measure generation adequacy risk for a particular generation background. This is particularly important for this project where the ability of a liberalised market framework to induce timely and sufficient generating capacity investment is of interest. A number of risk metrics typically used in power systems are formally defined and methods to calculate them explained.

2.8.1 Risk metrics

When developing models to determine loss of supply indices, it is important to consider the exact definitions of those indices. Listed below are some of the most common risk metrics used in the power sector.

1. *Loss-of-load probability* (LOLP) in period t is defined as the probability that the amount of available generation is unable to support a particular value of load:

$$LOLP_t = p(X_t < D_t) \tag{2.11}$$

   where X_t is the available generation and D_t is the system demand, both of which are random variables.

2. *Loss-of-load expectation* (LOLE) is the expected number of periods over a given time horizon, T, that available generation is unable to meet the load:

$$LOLE = \sum_{t}^{T} E\left[1_{X_t < D_t}\right] = \sum_{t}^{T} LOLP_t, \tag{2.12}$$

   where $LOLP_t$ is the LOLP in period t. A typical example is a time horizon of one year with periods of one hour. Note that different LOLEs can only be directly compared if both T and t are equal. For instance, if the time period is 1 hour then an outage of 2 hours at peak load would be recorded as 2 hours, in contrast if the time period is 1 day the same outage would effectively be recorded as being 24 hours long [61].

3. *Expected energy unserved* (EEU) is the expected volume of energy not met over a given time horizon. This risk index "*considers the severity of power shortages as well as*

their existence" [61]. Assuming no demand reductions can be applied, the EEU is given by $EEU_t = (D_t - X_t) \cdot LOLP_t$. Alternative equivalent metrics are Loss Of Energy Expectation and Expected Energy Not Supplied [61].

Other risk metrics, which are not formally defined here include *expected duration of load curtailment*, *average load curtailed*, *average energy not supplied* and *average duration of curtailment*. The majority of these measures relate to the risk of load curtailment (or load shedding). "Curtailment" is carried out in order to prevent system instability and, in severe cases, rolling blackouts. Usually the SO can take actions to reduce voltage and frequency, which may be enough to avoid a 'failure' event and the high economic and political costs associated with it. Consequently, the risk metric may be calculated based on the reduced load. This was the case in [62] where failure was defined as the requirement to disconnect load, assuming that load would first have been reduced by 7.5% via voltage and frequency adjustments [63].

Full mathematical details about how to calculate these measures can be found in [64]. In addition, [63] provides a comprehensive review of the use of these risk metrics in the GB system. This paper also looks at the influence risk modelling has had on the GB transmission system, both on operational and planning timescales.

2.8.2 Proxies for risk

Capacity margin can be viewed as a proxy for the level of security of supply risk. It is typically defined as the difference between total installed capacity (TIC) and most probable peak demand (MPPD), expressed as a percentage, i.e.,

$$CM = \frac{[TIC] - [MPPD]}{[TIC]}. \tag{2.13}$$

The time scale considered is important. For instance, *planning reserves* refer to the maintaining of system adequacy in order to meet demand in the long-term, whereas *operating reserves* are required to handle short-term issues, such as response to a disturbance [3]. The planning time scale is of interest here in identifying if enough capacity has been built in the first place. Fig. 2.18 demonstrates how planning and operational margins converge to real-time to produce "outturn" (or realised) margins.

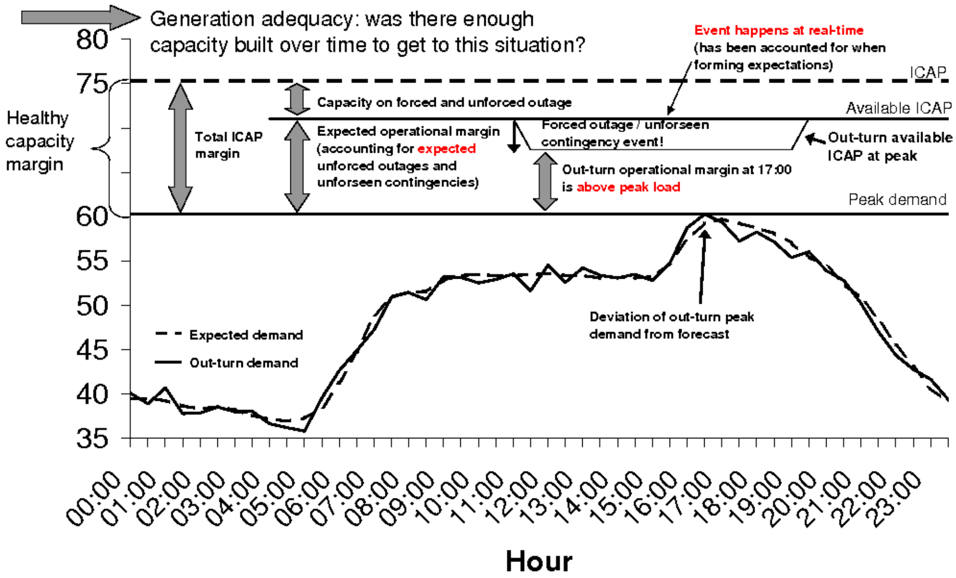


Figure 2.18: *Demonstration of how adequate investment in generation over time converges to operational time scales.*

With the introduction of high penetrations of variable generation with resource interdependence (i.e., output is correlated across different sites owing to the nature of the energy source), such as wind power, physical capacity margins will increase, however levels of generation adequacy risk will not necessarily move in the same direction. Therefore using the capacity margin as a proxy for system risk is no longer a credible approach.

A more recent concept used to analyse system security is de-rated capacity margin (e.g., [65] and [66]). The de-rated margin (DRM) is the ratio of de-rated capacity (DRC) to MPPD, expressed as a percentage:

$$DRM = \frac{[DRC]}{[MPPD]} - 1 \tag{2.14}$$

The DRC is computed by scaling installed capacity by expected availability at peak demand. The use of de-rated margin is preferable when calculation of an absolute level of risk is difficult, and it provides a robust alternative to the full capacity margin. Moreover it can easily be compared with the SO's estimate of what constitutes an acceptable margin (e.g., [66] and [67]). An example calculation of historic de-rated capacity margins in GB is given in Section 3.6.

2.8.3 Demand

From a risk modelling perspective, the available generation and demand at each time period are random variables. Demand patterns can usually be predicted well at the short operational timescale (i.e., hour-by-hour), though looking ahead brings more uncertainty and hence prediction is challenging. Factors such as weather, demand management, seasonal trends all contribute to the predictions. The main uncertainty on the operational timescale is weather variation. On a planning timescale, additional uncertainties about underlying long-term demand trends must also be considered [61]. For example, in GB some analysts forecast that increased demand owing to economic growth will be offset by increased energy efficiency [68].

2.8.4 Conventional generation

When considering conventional generation, the LOLP (2.11) can be calculated via the Capacity Outage Probability Table (COPT) method developed by Billinton and Allan [64] ([69] contains a neat explanation on implementation). Available capacity at each period (e.g., one hour) of a particular unit is a random variable, characterised by the unit's Forced Outage Rate (FOR). The FOR is the proportion of hours a unit is unavailable due to a forced outage, and to a good approximation FORs between different conventional generators are independent [61]. A common approach is to assume that units will be available with full capacity or on outage with zero available capacity. More precisely, if the capacity of unit u is c_u and its FOR is ρ_u then the expected available capacity is given by $E(G_u) = c_u(1 - \rho_u)$. That is, the distribution of the unit's available capacity follows a Bernoulli distribution between zero and c_u. So the unit is either available at full capacity with probability $1 - \rho_u$ or on full outage with with probability ρ_u. Methods to account for de-rated states are covered in [64], however the two-state approach is most common (e.g., [61, 70]).

Broadly speaking , the COPT involves a iterative calculation, adding each unit u to the model one at a time based on c_u and ρ_u. For instance, the iterative expression used to calculate the probability of a forced outage of X (e.g., MW) after unit c_u with FOR ρ_u is added to the system is given by [64]:

$$p(X) = p^{'}(X)(1 - \rho_u) + p^{'}(X - c_u)\rho_u, \tag{2.15}$$

where the primed values represent the probability of the forced outage of X before c_u is added. If m_u units of type n share the same capacity and FOR characteristics (e.g., this may be the case for units of the same fuel type) and are subject to independent forced outages then they can be treated as a single pseudo-unit with a distribution with the following moments:

$$E(G_n) \quad = \mu_n \quad = m_u c_u (1 - \rho_u), \tag{2.16}$$

$$Var(G_n) \quad = \sigma_n^2 \quad = m_u c_u^2 \rho_u (1 - \rho_u), \tag{2.17}$$

where each (pseudo-) generator, G_n, has capacity $c_n = m_u c_u$. This is a useful property for a system with a large number of units and can lead to a reduction in computational intensity when calculating the COPT. A demonstration of how systematic deviations in the FORs for a group of 'similar' units can lead to significant differences in the overall availability distribution of conventional generation (and the risk metric) is given in [71]. This paper also includes a novel technique to model ρ_u as a random variable when calculating the distribution and determines the error bars on the system LOLP.

The FOR encapsulates forced (or unplanned) unit outages. Units also undergo planned maintenance outages throughout the year, which must also be considered when calculating year-round system risk. The nature of a planned outage is that the SO is given adequate warning of when they will occur and therefore has time (at the operational level) to plan the system accordingly. There may also be an option to shift outage schedules when margins are tight [61]. Fortunately during periods of highest demand (when wholesale prices are typically highest), most generators will try to make available as much capacity as possible. It is argued in [61], that under the assumption that these peak demand hours dominate the risk calculation (i.e., the LOLP (2.11) is highest in these hours so contributes significantly to the year-round LOLE (2.12)), it might be reasonable to ignore planned outages, and assume that the distribution for the availability of units is the same in all hours of the year.

2.8.5 Wind generation

One of the main concerns when considering wind generation is its variable nature and the impact on system reliability. Even if countries install many wind turbines, the amount of wind generation will seldom reach total installed capacity. For instance, wake effects, where output is reduced for turbines situated behind the first row, are a feature of large farms. In addition, wind farms are constructed so the turbines harvest the wind from a certain direction and they lose efficiency in varying wind direction. Also due to spacial variation in the resource, when one farm is active another may be inactive. By installing more wind capacity, the chances of harvesting more wind is increased.

The *capacity credit* (CC), or *capacity value* (used interchangeably), is common when measuring the contribution of renewable energy generation to meeting demand. CC is not a new concept in power systems and in fact has been used for many years in relation to conventional sources of generation. A number of different definitions of capacity credit exist in the literature, most of which are dependent on the method of calculation. Essentially they all concern the ability of plant to support load and are independent of the risk metric used. A discussion of the various definitions can be found in [61] and [72]. In [61] the definition used is the *effective load carrying capability* (ELCC), which can be summarised as [61]:

1. Calculate the risk index before the additional generation is introduced.
2. Introduce the extra generation and re-calculate the risk index.
3. The ELCC of the extra generation is the additional demand which returns the risk index to its original value.

That is, the ELCC for a particular level of additional generating capacity estimates the amount of additional demand that can be served due to the extra generation whilst maintaining the original level of system risk [72].

Another available method is the *effective firm capacity* (EFC), which is calculated as follows:

1. Calculate the risk index before the additional generation is introduced.
2. Calculate the risk with the additional generation introduced.
3. The EFC is the amount of firm capacity whose addition gives the same reduction in risk as addition of the new generation.

There is virtually no difference between EFC and ELCC for small penetrations of variable output generation. In contrast for larger penetrations the difference is noticeable. For instance, consider two wind distributions for a large installed wind capacity; they are similar for low wind availabilities, but the first has a much higher probability of very high wind availability. This first distribution will give higher ELCC, but the EFCs will be very similar as there is basically no risk when wind availability is high, so the precise detail of that part of the distribution does not influence the result of EFC [73].

The CC is typically expressed as a percentage; CC = [ELCC or EFC]/[total extra generation]. The CC of wind is an important metric in this type of work because it provides a measure of wind's contribution to overall system reliability, and it can be used to de-rate wind capacity when measuring the de-rated capacity margin.

2.9 Chapter summary

This chapter has covered some of the fundamentals of power systems economics. In Section 2.2 an overview of 'energy-only' markets and ancillary services was provided. Next in Section 2.3 some key economic concepts were introduced. The key findings of this review were the calculation of generator short-run marginal cost and how, under a perfectly competitive market, the last generator dispatched in the economic dispatch provides the system marginal price. Also the concepts of scarcity rent and wealth transfer via price mark-ups in imperfect markets formed a vital part of the review. In Section 2.4 attention turned to long-run issues, in particular generation investment and how as supply tightens or demand expands, price spikes increase scarcity rents, which is viewed as an investment signal to generators. In section 2.4.3, the minimum cost of supply was derived using load duration, technology screening and price duration curves. This demonstrated that both energy and capacity prices are required in order to determine the minimum cost of supply. In Section 2.5 the cost of generation was explored, and particular attention paid to technologies used in later investment modelling. Following this, Section 2.7 gave an overview of capacity mechanisms available to market designers aimed at mitigating the generation adequacy problem. Finally, Section 2.8 surveyed methods of measuring generation adequacy risk in power systems via metrics such as LOLP and LOLE. Also a discussion on the use of de-rated margins as a proxy for risk, particularly in system with high penetrations of wind power, was included.

Chapter 3
Experience in GB

As this thesis is concerned with the GB energy market, it is important give a brief overview of the ESI to date. This brief history of the ESI illustrates how security of supply concerns have been addressed to date. This is followed by an account of liberalisation including an overview of market structures. Finally, some of the key modelling works applied to the GB market in recent years are reviewed.

3.1 Pre-liberalisation

The ESI in Britain was initiated through The Electricity Lighting Act 1882 and a 132kV National Grid began interconnected regional grid operation in 1933. This review begins with the creation of the centrally managed system some years after the Electricity Act 1947.

Shortly after the end of the second World War, the Electricity Act 1947 was passed and the numerous (well over 600) small-scale local electricity suppliers were merged into sixteen larger firms. During this time, consumers were forced to buy their electricity from the monopoly utility operating in their area. In 1957 the Central Electricity Generating Board (CEGB) was created and made responsible for electricity generation in England and Wales (E&W) with two similar entities in Scotland; the South of Scotland Electricity Board (SSEB) (circa 1955) and the North of Scotland Hydro-Electric Board (NSHEB) (circa 1943). These companies owned the generation and transmission assets in their respective areas and were responsible for the security and quality of supply. The distribution network in E&W was made up of 12 area electricity boards (AEB) who were responsible for the distribution of electricity to end consumers (i.e., also monopolies but within a smaller region). In the case of the Scottish network, these were fully vertically integrated monopolies with both the regional transmission and distribution grids managed by SSEB and NSHEB. These structures are summarised in Fig. 3.1. As the name suggests, the CEGB was a central organisation with generation expansion planning driven by cost minimisation. It was a classic example of a cost-of-service regulated public utility and its generation fleet was dominated by coal and (eventually) nuclear plant [74].

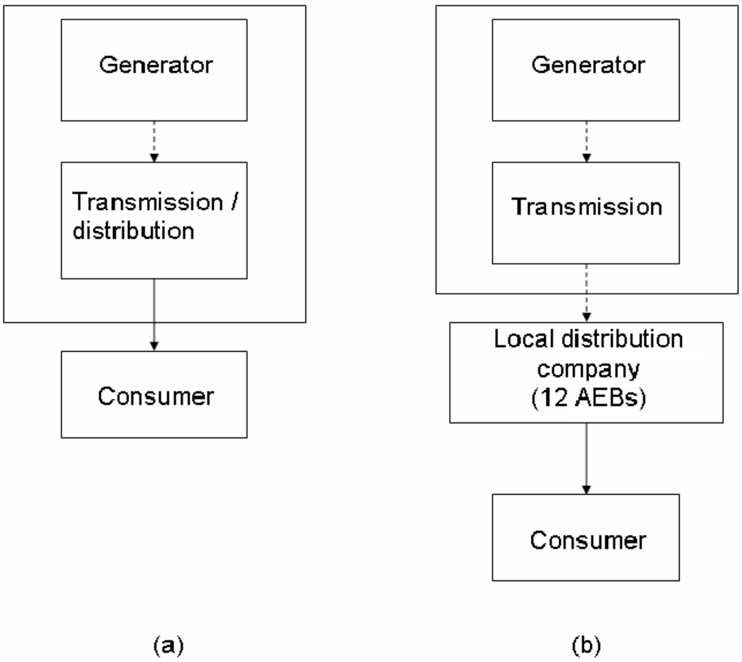


Figure 3.1: *Industry structure under (a) the SSEB and NSHEB and (b) mildly less monopolised CEGB. Inspired by [5].*

3.1.1 The security of supply standard

To tackle the challenge of generation adequacy, standards were developed and any additional generating capacity required was constructed at the optimal time. With it taking around 2 years to build a gas turbine (GT), a planning horizon of 7 years was regarded as adequate to address any upcoming generation shortfall in good time. It was decided that a target risk level of 3% winter peak LOLP, which translated into a 28% planning margin was adequate [75]. These target levels were later changed on cost-benefit grounds to 9% and 24% respectively [62]. The cost-benefit approach calculated the LOLP under the following assumptions:

- Frequency and voltage reductions can reduce demand by up to 7.5%.
- Cost of providing incremental capacity for the planning margin of about £15/kW p.a. at March 1982 prices [75].

- A VOLL of £2/kWh.

The recommendation to relax the standards to 9% was made on the basis that the incremental capacity costs would rise to £30/kW per annum owing to the increased costs surrounding new high cost coal-fired plant, which was the technology of choice at the time (there was some PS and hydro capacity in Wales with nuclear plant being deveploped latterly). The other justifications included the VOLL figure dropping to £1/kWh together with additional assumptions surrounding disconnection minimisation.

3.2 Liberalisation

In 1990, under the third parliament of a Conservative government the CEGB was restructured and privatised. Deregulated markets had already been implemented in supply of gas, airlines and transportation and it was considered time to do the same with the electricity industry [5]. The CEGB was divided between four companies; the high voltage transmission network became the responsibility of the National Grid Company (NGC), a not fully independent company. The generation fleet was divided between National Power (50%), PowerGen (30%) and Nuclear Electric (NE), with the PS capacity at Dinorwig and Ffestiniog going to NGC [76]. NE consisted exclusively of the UK's nuclear power capacity and was kept under state ownership as it was felt the nuclear reactors were too expensive to be privatised [76]. NGC was not fully privatised until 1996 (floated on London Stock Exchange on 11 December 1995) because there were no other privatised transmission companies to use as a comparison to obtain a fair share price value. Up until its privatisation, NGC was owned by the 12 AEBs. It was around the time of NGC privatisation that it's PS generating capacity was sold to the American firm Edison Mission Energy (acquired by International Power in 2004) as it was deemed anti-competitive to own both generation and transmission assets.

3.2.1 The England and Wales Power Pool

Much has been written about the history of liberalisation in GB and none more so than the period immediately after industry privatisation during the days of the Pool system. This section contains a brief overview of that period, but is by no means comprehensive. A particularly informative account of the history of the pool can be found in [77] or for an analysis of its competitiveness see [78].

The E&W Pool went live on 1 April 1990. It was based on a classic pool system, whereby the 48 half-hourly prices for each day were determined by generators submitting a supply curve for each period at the day-ahead stage. The SO then computed the least-cost economic dispatch required to meet demand in each period and all committed generators were paid the SMP.

As real-time approached, the differences between actual and forecast demand, and generation, together with transmission constraints would usually lead to the day-ahead dispatch being adjusted. The additional costs incurred by the SO (i.e., ancillary services) were covered via a price *uplift* element. The breakdown of this *uplift*, together with the availability payment, which was paid to generators dispatched at the day-ahead stage but were not dispatched at real-time, can be found in [79]. In addition to trading on the spot market, financial swaps and Contracts for Differences (CfDs) were facilitated. CfDs allowed participants to exchange volatile pool prices for a fixed price stipulated in the contract [7].

3.2.1.1 Market for capacity

In addition to the spot market for energy, the Pool included a market for capacity. This comprised a capacity payment which was added to the SMP at each half-hour period to give the Pool purchase price (PPP). This additional payment was centrally administered and was a function of the system LOLP and VOLL:

$$PPP = SMP + (LOLP \cdot (VOLL - SMP)). \tag{3.1}$$

where VOLL was set at £2000/MWh in 1990.[1] This was increased in the GB regulation model of 1999 to £2770/MWh.[2] A plot of the system LOLP and PPP for the financial year 2000/01 is presented in Fig 3.2. The plot clearly shows how the PPP rises sharply in periods where the LOLP is highest, which interestingly do not always coincide with periods of highest demand.[3] Note that the uplift element was then added to the PPP to give the Pool selling price (PSP).

[1] Likely as a result of existing CEGB standards.

[2] This VOLL was derived from a Finnish study from the 1970s; it was converted into pounds sterling and corrected for inflation [80].

[3] For a commentary on some of the more interesting price movements under the pool (inc. underlying reasons), see p.22 of [81].

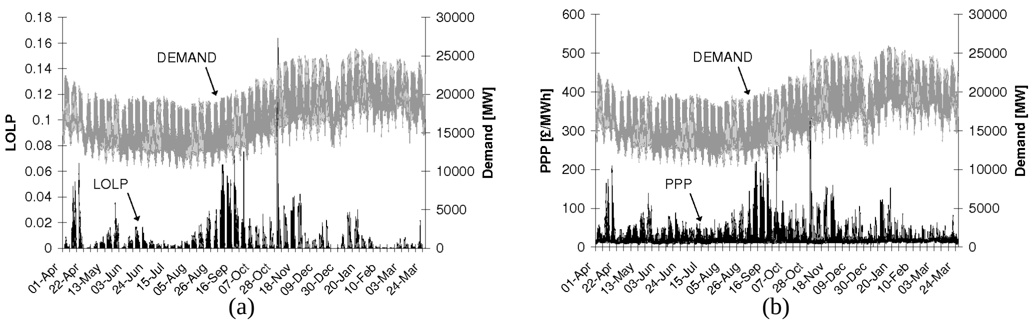


Figure 3.2: *(a) Plot of system LOLP and (b) PPP with E&W demand. Source: [82].*

The capacity payment was designed to ensure that generators received the revenues needed to invest in new capacity. It was engineered to increase exponentially as demand approached total available system capacity [79]. As pointed out in [77], producers with a large market share can exploit a capacity payment that is a function of system LOLP by declaring plant as unavailable - an action which lowers the available reserve, thus increasing the LOLP and capacity payment element of the wholesale price.

3.2.1.2 Issues with the Pool

Owing to their substantial market share, and ability to exercise market power, National Power and PowerGen set the spot price over 90% of the time although they supplied less than 60% of total electricity generated [77]. Observations of this type led to an investigation by the regulator, OFFER; it was concluded that after only a few years of operation, Pool prices had risen from a lower than expected level in the first year to a level that was unacceptable. Consequently, the large incumbents began to divest their generation assests, the most notible of these was the sale of 6 GW of capacity to Eastern Electricity during 1994/5 [76]. In addition a temporary price cap was in place between 1994-96 while the sales were finalised [74].

Toward the end of the 1990s, electricity generation had been restructured from a market dominated by three firms to one where the generation market share by a single participant was no higher than 20%. In addition to heavier regulation, the shift can be attributed to the fall in demand for coal and a ‘dash for gas’. This prompted a huge influx of CCGT entry into the generation market from a number of new IPPs. This gave rise to another problem surrounding interactions with the UK gas market. Electricity prices were set at the day-ahead stage while the gas market operated much closer to real time. Thus, gas generators could influence electricity prices set at the day-ahead stage, then if prices were unfavourable, they could sell their gas closer to real time on the gas market, adding further complexities to the issue of short-term supply-demand matching [83]. Such was the discontent with the pool system that a review of electricity trading arrangements was ordered. In 1998, the review [84] recommended that the pool system be replaced by the New Electricity Trading Arrangements (NETA), which went live in March 2001.

3.2.2 NETA and BETTA

The British Electricity Trading and Transmission Arrangements (BETTA) were established in April 2005 when the electricity market in E&W merged with its counterpart in Scotland into a single GB wholesale market.[4] Prior to then, the market in E&W operated under NETA.

3.2.2.1 Structure

There are currently over 40 generation firms in GB, yet 70% of generation is owned by the so-called 'big six' generators: Scottish and Southern Energy (SSE), RWE Npower, EdF, E.ON, Centrica and Iberdrola/Scottish Power (SP). Many of these entities are vertically integrated, participating in both the generation and retail side of the market. The two Scottish firms also own transmission assets in Scotland with most of them also owning one or more of the 12 licensed distribution operators. To ensure that trading between parties who have interests in more than one element of the market provides a cost effective service for consumers and facilitates competition, market regulation is required. Under BETTA this role is carried out by OFGEM.

3.2.2.2 Trading arrangements

BETTA facilitates bilateral trades between generators, suppliers, traders and customers on a rolling half-hourly basis [27], depicted in Fig. 3.3. These bilateral trades can divided into two categories [85]:

[4] Northern Ireland has been participating in the Single Electricity Market in Ireland since 2007.

- Forwards and futures contract market (or structured contracts): allows electricity trading between generators and suppliers over longer periods (a year or more) without formal price disclosure. In GB over 90% of electricity is traded in this way.[5]

- Power Exchanges: allow parties to trade a range of electricity products including half-hourly, peak and base-load and day-ahead contracts [87]. Examples in GB include the APX Group and the Intercontinental Exchange. About 3% of electricity is traded in this way. These exchanges also publish price information which can be used as reference for over-the-counter platforms and structured contracts.

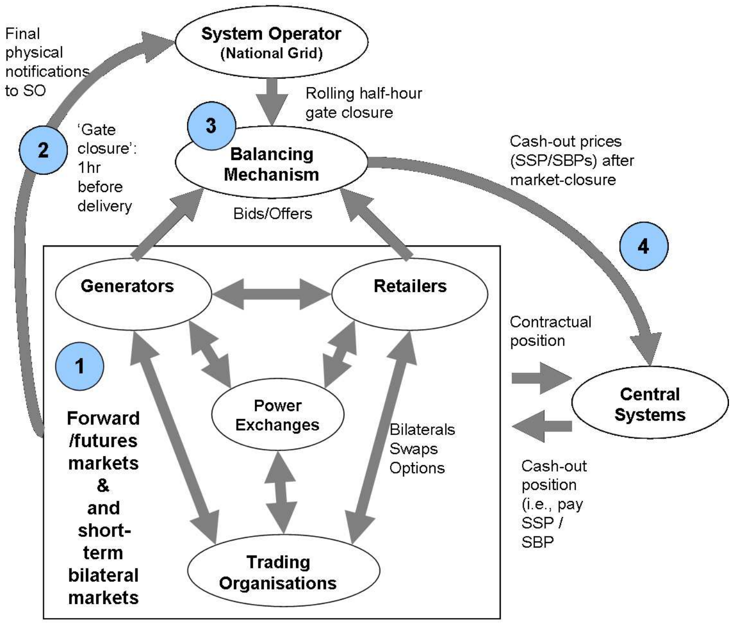


Figure 3.3: *Overview of financial and physical flows under NETA/BETTA with numbers to indicate order of events. Based on diagram in [7].*

On top of these bilateral mechanisms there is also a balancing process, known as the Balancing Mechanism (BM), which is operated from gate closure (set at 1 hour) by the SO in order to ensure supply matches demand. Through this mechanism, participants can submit *bids* to decrease generation (or increase demand) and *offers* to increase generation (or reduce demand). The mechanism operates on a 'pay as bid' basis. Some 2-3% of electricity is traded in this way

[5] Interestingly, the regulator, OFGEM, is in the process of undertaking a review on GB wholesale market liquidity and has launched a consultation as part of the process [8]. Broadly the responses state that forward market liquidity should be deeper, and that it is difficult to hedge a long-term position. There is also concern that market trading for small players could be better facilitated [86].

[85]. Generators who are requested to deviate from their contracted obligations as a result of system imbalances (e.g., network congestion), will still receive their contracted payments and save on fuel not burnt.

Any participants who stray from their 'Final Physical Notifications' (FPNs) must pay one of two prices. Those whose generate more (resp., consume less) will pay the System Sell Price (SSP). If, for whatever reason, a generator falls short of its contracted volume, or a retailer requires additional energy, then they must pay the System Buy Price (SBP). Both these prices are higher than the prices stipulated in the bilateral agreements due to the restricted time available to deal with this imbalance problem. There is no price cap under BETTA, although the BM can only enter four digits, making the maximum allowed price £9999 [2]. The revenue received from SBPs and SSPs is used to cover the additional costs imposed on the SO as a result of having to balance the system.

3.2.2.3 Ancillary markets

Additional revenue may be obtained from provision of ancillary services, although agreeing to provide these services, means that participation in the energy market is restricted or forbidden. These products are procured by the SO. Of particular interest is the short-term operating reserve (STOR) market, which is a targeted 'pay-as-bid' tender for capacity. Generators submit bids for unit availability and utilisation to the tender and participation in the energy market is restricted. It is targeted because the technical characteristics of plant must also be submitted. The tender includes two types of contract; one for flexible and one for committed capacity. Flexible contracts are from 1 or 2 seasons up to 2 years ahead. And committed contracts are from 1 or 2 seasons up to 15 years ahead. *Flexible* units declare themselves available within certain time windows (Fig. 3.4); these windows are not fixed and can be altered up until a week ahead. These units do not participate in the BM. *Committed* units must be available within a window and can be BM or non-BM units. Payments for availability are withheld if the unit is unavailable during an agreed window, thus encouraging a good reliability record.

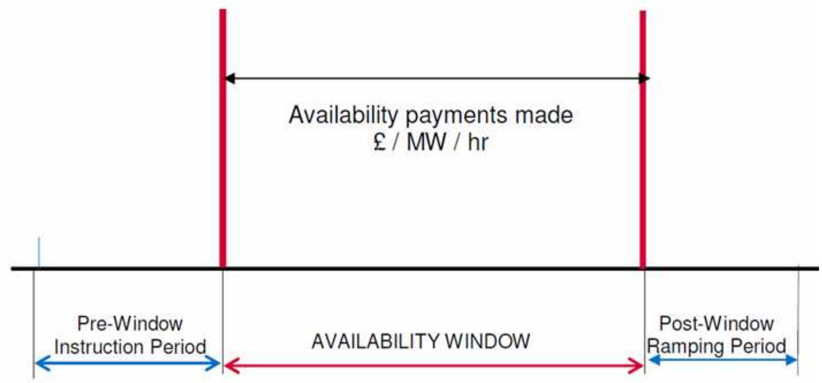


Figure 3.4: *Availability windows and payments under STOR. Imported from [88].*

3.2.2.4 Summary

Since their introduction, it is generally accepted that NETA/BETTA has successfully enabled competition to fully emerge in the generation market for the first time [7]. The purpose of these new trading arrangements was to prohibit generators with substantial market power from abusing their position and adversely affecting consumers or distorting competition between companies. The debate about whether this was a consequence of the new trading arrangements, restructuring, or a combination of both, continues.

3.3 Capacity margin oscillations in GB

The historic capacity margin for GB at realised peak demand is plotted in Fig 3.5 (solid line). Also plotted is the theoretical capacity margin at the year-ahead stage using the NG forecast (dashed line); this data was only available back to winter 2001/02 hence the shorter period. This is included to provide a better illustration of the perceived generation adequacy risk at the planning stage. To illustrate the influence of analysing alternative (but deemed equally reliable) data sources, this is compare with the margin used in [20] (dotted line), which looked for evidence of capacity cycles. This data was taken from a UK Parliament note on security of supply [89] and plainly gives different values of capacity margin. They differ due to alternative sources of installed capacity and peak demand data used when constructing the margin. That said, both lines exhibit an oscillating pattern overall (as confirmed by analysis in [20]), with an observable increasing trend in later years. The signal shows symptoms of a sustained oscillation around 20-25%.

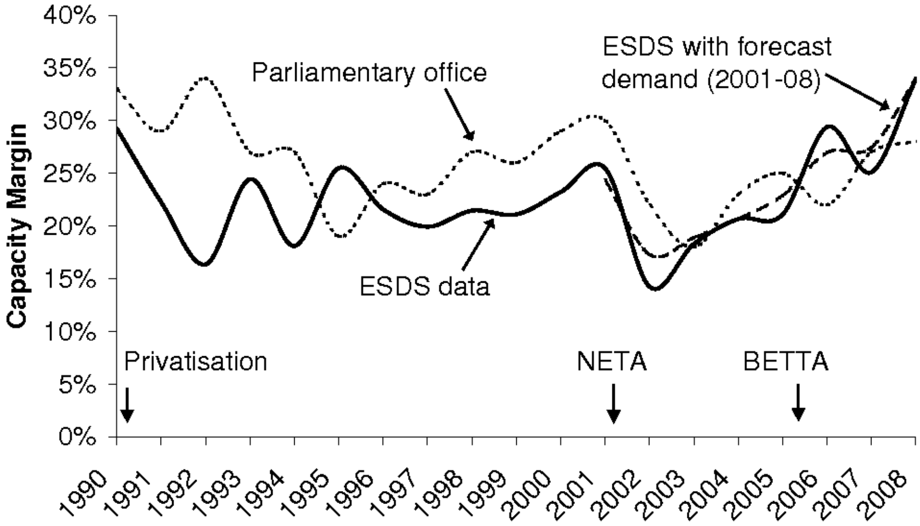


Figure 3.5: *Generation capacity margin in the GB since industry privatisation. The solid line uses data from the Economic and Social Data service (ESDS) [90] and the dotted line uses data from [89]. The shortened dashed line is ESDS data combined with NG forecast peak demand [73]. Dates of significant market framework changes also shown (indicated with arrows to x-axis).*

3.4 Generation adequacy in GB

The risk metrics described in sub-section 2.8.1 have been employed in GB system operation and planning in the past. Firstly, a LOLP (2.11) calculation was carried out for the CEGB standards. Secondly, the method for capacity pricing presented in [70] was subsequently used to inform the capacity payment system under the E&W Pool. Finally, the GB SO issues a annual outlook statement containing the perceived levels of system risk for the coming winter and summer. The published availability of the conventional GB generation fleet for the winter 2010/11 [66] is summarised in Table 3.1.

It has been widely accepted in recent years that a benchmark capacity margin at or above 20% for a predomiately thermal system provides an acceptable level of risk. Given that peak demand is typically around 60 GW in GB (Fig 3.6), Table 3.1 suggests this current system remains broadly inline with this benchmark ($100\% \cdot (74.4 - 60)/74.4 = 19\%$). However for the reasons discussed earlier, a 20% capacity margin is no longer adequate once high volumes of variable

Power station type	No. units	Full metered capacity (GW)	Assumed availability	De-rated capacity
Nuclear	22	10.1	0.75	7.6
IFA	1	2	1	2
Hydro	9	1.1	0.6	0.7
Wind	-	2.5	0.1	0.3
Coal	62	27.9	0.9	25.1
Oil	4	2.7	0.8	2.2
Pumped storage	16	2.7	1	2.7
OCGT	34	1.2	0.9	1.1
CCGT	124	27.5	0.9	24.8
	272	**74.4**		**66.3**

Table 3.1: *Summary of transmission connected conventional unit types for 2010/11 [66] .*

output non-dispatchable generation is introduced.

3.5 An historic generation adequacy risk calculation

When making a forecast about future generation investment trends and resulting levels of generation adequacy, it is helpful to consider historic levels of system adequacy risk. Further, any debate about whether or not a capacity mechanism is required must be informed by a robust generation adequacy risk measure. This permits assessment of the risk in any given investment scenario, and also guides the design of energy and capacity markets by setting a baseline for what the market should deliver (e.g., an adequate capacity mix should be consistent with appropriate returns on investment) [91]. Further, by carrying out a historic generation adequacy risk calculation, one can gain an insight into how relative levels of risk have evolved in recent years and what levels of relative change have been experienced.

These results will also inform the debate about how the emerging trend of reduced capacity margins in GB since market liberalisation have impacted security of supply. For instance, weather-adjusted historic capacity margin data presented in [12] for E&W, Finland, Norway, Sweden and the US (national average) shows a general trend of a reduction in GB capacity margins since market liberalisation at the start of the 1990s. This pattern is evident in Fig. 3.5 for the period immediately following industry privatisation. This has increased the overall efficiency of the system (holding too much capacity is expensive [12]), but by looking at relative levels of system risk we can go some way to determining the impact on security of supply.

3.5.1 Historic demand time series

Historic half-hourly demand data from April 2001 (i.e., NETA introduction) is available from the GB SO, National Grid (NG) from [19]. Such is the climate in GB that the highest demands are driven by low temperatures and occur during winter (November-March). It is during these high demand hours that the adequacy risk is typically highest. With this in mind, the historic risk calculation has been carried out for the demand hours spanning 10 winters 2001/2 - 2010/11. The GB 'IO14_DEM' data is the most applicable for generation adequacy calculations because this is based on operational generation metering and includes station load and pumped storage (PS) pumping [61]. However the GB 'IO14_DEM' data is inconsistent; before April 2005 it is for England and Wales only. 'INDO' demand, which excludes station load and PS pumping, is available for the entire period and the winter 'IO14_DEM' can be approximated by 'INDO' plus 600 MW. This is the approach taken here when 'IO14_DEM' data is not available.

To account for underlying changes in both demand patterns and the absolute levels of peak demand, each winter period's demand is normalised by its realised 'Average Cold Spell' (ACS) winter peak (ACSWP) demand and rescaled to 60 GW:

$$d^* = \frac{60 \cdot d}{ACSWP} \tag{3.2}$$

where d is the realised demand and d^* is the result of the rescaling. These values are displayed in Table 3.2. ACS peak demand is forecast each year in advance of the forthcoming winter by the SO, and is described as having *"a 50% chance of being exceeded as a result of weather variation alone"* [92]. The realised ACS peak is calculated post winter and is a measure of what peak demand would be given a winter's underlying demand patterns and "typical" winter peak weather conditions [93]. This makes it suitable value for the normalisation. These values are plotted in Fig 3.6. Demand data is been transformed to an hourly resolution by taking the hourly demand to be the maximum of the two half-hour periods.

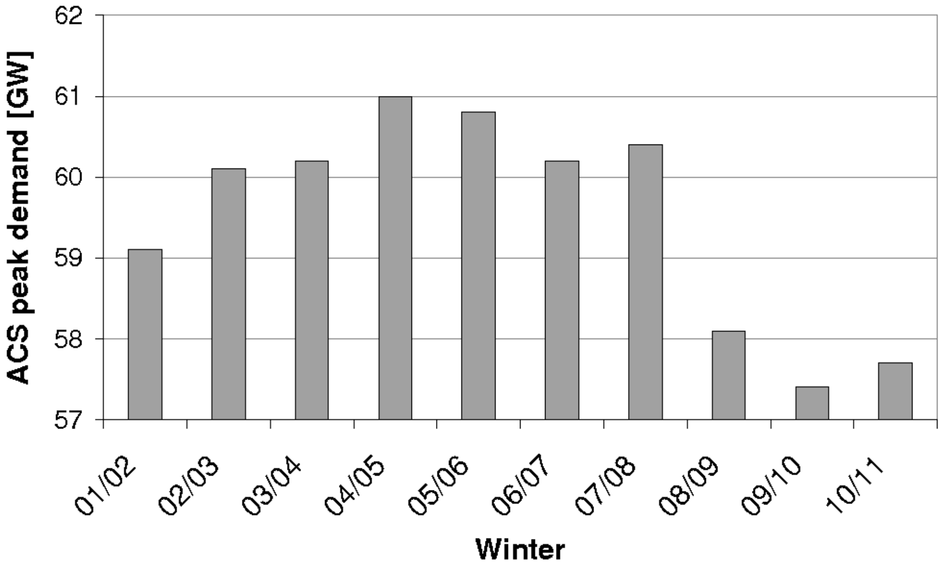


Figure 3.6: *Plot of ACS peak demand in GB for winters 2001/02 - 2010/11.*

	01/02	02/03	03/04	04/05	05/06	06/07	07/08	08/09	09/10	10/11
ACS peak (GW)	59.1	60.1	60.2	61.0	60.8	60.2	60.4	58.1	57.4	57.7
Norm. peak	58.6	60.6	59.3	58.2	59.2	57.9	60.2	61.1	61.7	62.1
De-rated capacity	66.2	63.44	64.4	66.3	68.1	65.6	65.3	64.9	66.1	66.3
De-rated margin (%)	11.9	5.6	7.0	8.6	12	9.0	8.1	11.7	15.2	14.9
Norm. de-rated (%)	12.8	4.7	8.7	13.8	15.1	13.4	8.5	6.3	7.1	6.8

Table 3.2: *Data for de-rated capacity used in Fig. 3.9. Data on de-rated capacity is contained in NG Winter Outlook reports 2005/06 - 10/11, e.g., [66]. De-rated capacity for winters 2001/02 - 04/05 estimated by de-rating total installed capacity (using ESDS data [90]) by 0.9.*

3.5.2 Probability distribution for available conventional generation

The next step is to construct a probability distribution for available conventional generation. Here, the term conventional generation covers all forms of generation currently connected to the high voltage transmission system in GB, bar wind. Technical plant availability data is not available in GB. The approach here is to use generation unit data from the NG Seven Year Statement [27] for unit capacities and expected winter peak availability assumed in NG's Winter Outlook [66] as FORs. This data is summarised in Table 3.1 and is assumed reliable (given that NG use it to assess the generation adequacy risk for the forthcoming winter). The Unit Effective Capacity (UEC) in [27] has been used for all units, except those which are behind a transmission constraint (e.g., Peterhead), in this case the individual UEC are scaled so that total export capability is equal to the transmission limit. This resulted in the amount of available CCGT capacity in Table 3.1 reducing from 27.5 GW to 26.7 GW. Hydro units belonging to the same hydro scheme are combined into single pseudo-units owing to their dependence in terms of resource availability. These pseudo-units are displayed in Table 3.3. Using the COPT technique (2.15) with a 1 MW bin size, the probability distribution for available capacity is determined. This was implemented in the AIMMS programming language; the main procedure is provided in Appendix A.1. The aggregate probability density function (pdf) and cumulative distribution function (cdf) for available generation are displayed in Fig. 3.7. The mean of the distribution is 65.25 GW and standard deviation 1.80 GW.

Hydro scheme	**Capacity (MW)**
Affric Beauly	235
Breadalbane	90
Conon / Shin	136.3
Great Glen	110
Kinlochleven	30
Sloy / Awe	233
Tongland	33
Tummel Valley	229
TOTAL	1096.3

Table 3.3: *Combined hydro scheme pseudo-units. Based on [94].*

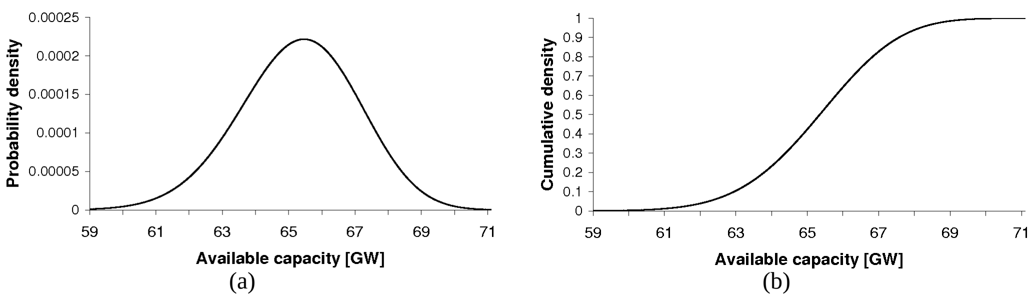


Figure 3.7: *(a) Plot of pdf and (b) cdf for available conventional generation calculated using COPT technique.*

Given that the GB SO's current view is that a de-rated margin of 5 GW capacity is required going into winter [66], this distribution seems reasonable, which is encouraging given the uncertainty associated with generator FORs. Furthermore, if one wanted to carry out a risk calculation for a "more risky" system (e.g., in the absence of closed LCPD plant), then it would be reasonable to shift the mean of the distribution and rescale the standard deviation (scales with the square root of the number of units) rather than rerun the calculation with some volume of plant removed.[6] This will also resolve all issues of exactly how to deal with transmission-constrained plant and emission constraints [73]. One such shift where 4 GW of capacity (8 units) was removed is shown in Fig. 3.8.

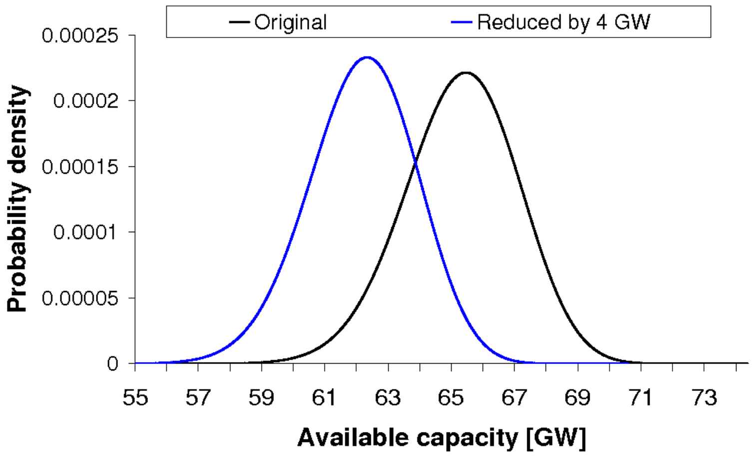


Figure 3.8: *Plot showing distribution for available conventional generation when 4 GW of capacity is removed.*

Using the first distribution shown in Fig. 3.7, the hourly winter LOLPs (2.11) can be computed. That is, LOLP $= F(d) = p(D < d)$ where d is the demand for that hour and $F(d)$ is the cdf for available generation (Fig. 3.7(b)). Hourly LOLPs can then be summed to produce the 10-winter hindcast LOLE (2.12). For simplicity, the probability distribution for conventional generation used to calculate the LOLE for each winter period remains unchanged throughout the 10-year analysis. The mean of 65.25 GW is broadly in line with the mean de-rated capacity for the whole period (65.65 GW) (Table 3.2), and given that the underlying mix and amount of capacity has not altered significantly over the period analysed, this approach is reasonable.

3.5.3 Treatment of wind

Given that the installed wind capacity in GB is currently quite low relative to the overall total installed capacity (3%, Table 3.1), the approach taken here is to de-rate wind and treat it as a pseudo-thermal unit in the COPT calculation. This is a reasonable approach given that 300 MW

[6]An algorithm for removing a unit from the original COPT without re-running the entire COPT calculation is presented in Chapter 2 of [64], yet this shift is an efficient option when considering the removal of multiple units.

of additional de-rated capacity is unlikely to affect the risk calculation significantly. Techniques to properly account for high penetrations of wind in the risk calculation are covered in Chapter 6.

3.6 Historic generation adequacy risk calculation results

Results of the 10 winter LOLE calculation are shown in Fig. 3.9. The average LOLE over the 10 winter period was 0.06 hrs/yr (or 6 hours in 100 years) with a standard deviation of 0.01 hrs/yr. The graph shows how the relative risk remains largely unchanged over the first 8 winters but rises sharply in the last two winters. The red LOLE bars indicate winters where realised peak demand exceeded the ACS peak forecast (which is in fact in nearly 50% of the years analysed). On closer inspection of the data, which is presented in Fig. 3.10, nearly 80% of the 10-year LOLE is contained in just 40 demand hours, with 15% contained in just 2 demand hours.

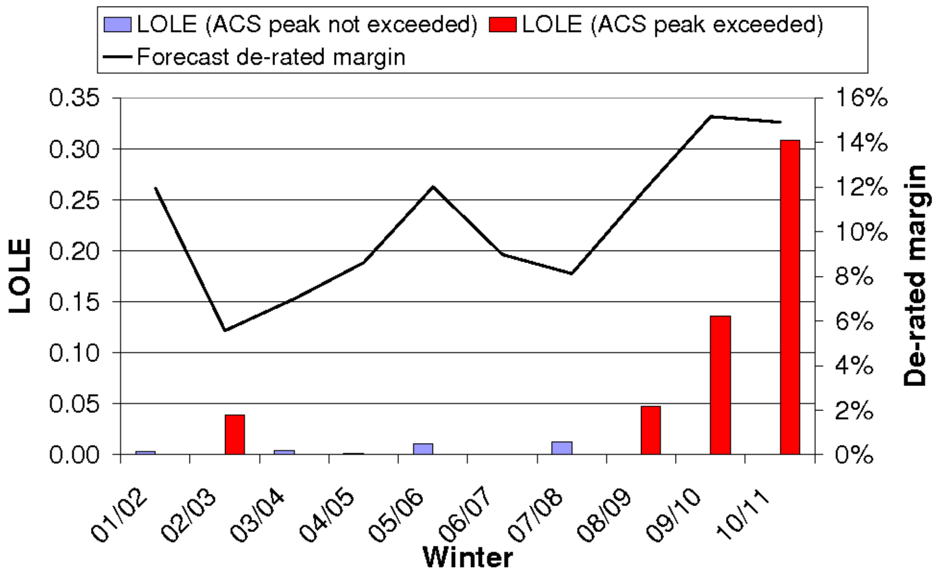


Figure 3.9: *Plot of LOLE results for 10 winters (bars). Also shown is forecast theoretical de-rated margin (black line), normalised hindcast de-rated margin (blue line), and winters where ACS peak was exceeded (red bars).*

Also shown in Fig. 3.9 is the historical theoretical de-rated capacity margin (black line). This is the de-rated margin that would have been forecast at the year-ahead stage given the de-rated capacity and forecast ACS peak published by NG. For instance, in 2010/11, it is the sum of the

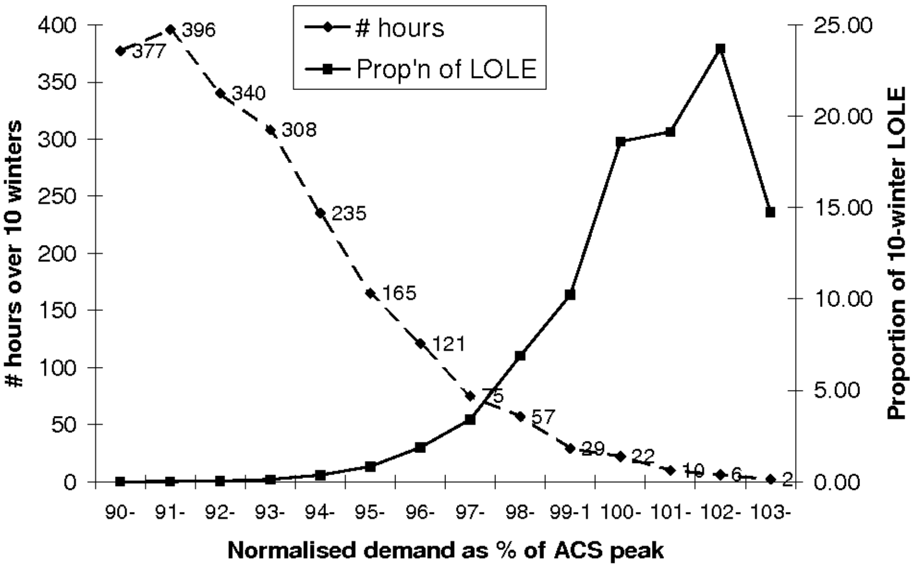


Figure 3.10: *Plot showing number of hours demand contained in LOLE (dashed line with hours marked) and proportion of LOLE in each 1% bin (solid line). The bins (x-axis) are the normalised load as a percentage of the 60 GW peak demand.*

fifth column in Table 3.1. Plainly the de-rated margin in the winters 2009/10 and 2010/11 is higher relative to historic levels whereas the hindcast relative risk calculation was in fact highest in these winters. This can be explained by the fact that realised demand was high relative to underlying demand levels and “typical” winter weather conditions. Or put another way, although levels of de-rated margin have been higher in recent winters relative to historic levels, a succession of mild winters up to 2009/10 has kept relative levels of generation adequacy risk down. The normalised de-rated capacity margin in Fig. 3.9 is calculated using the normalised realised peak demand figures (3.2). This acts as a proxy for the LOLE, and provides a picture of realised de-rated margins, compared to the forecast. The winter with the highest risk, 2010/11, demonstrates that even seemingly healthy margins can be eroded in the face of ‘high’ (but not extreme) peak demand; in this case reduction of 8% is witnessed.

The importance of considering relative levels of risk in this type of calculation is demonstrated when the underlying distribution for available conventional capacity changes. Using the probability distribution for the reduced generating fleet described by Fig. 3.8, the LOLE calculation was re-run. The results are shown in Fig. 3.11. The relative levels of LOLE remain unchanged, by comparison absolute values increase by an order of magnitude (the previous values of LOLE are barely visible at this scale). Note that the absolute levels of risk could also be altered by changing the normalising factor of 60 GW in (3.2), though the same relative levels of risk will be produced.

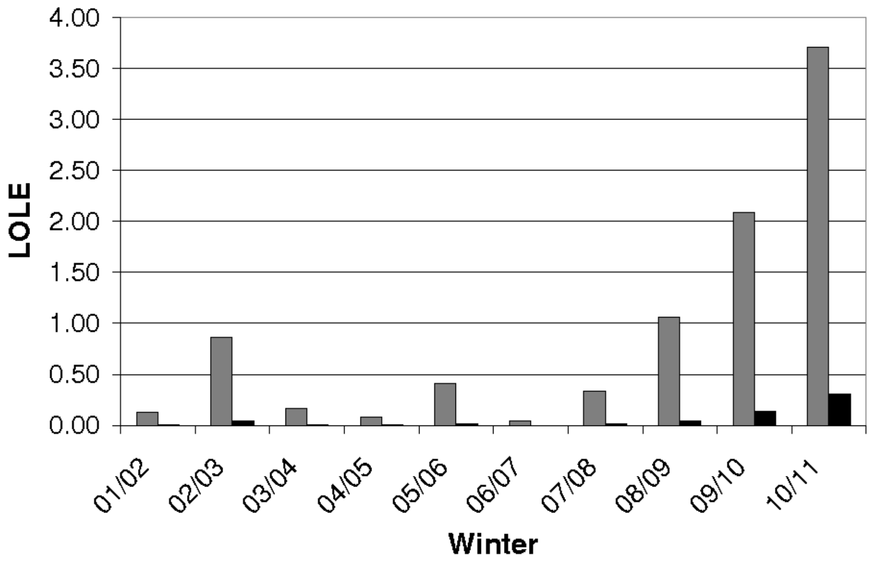


Figure 3.11: *Plot of LOLE results for 10 winters using reduced conventional fleet (grey bars). Also shown are LOLE results from Fig. 3.9 (black bars).*

3.6.1 Summary

The results of this historic risk calculation have demonstrated how relative levels of risk have changed in recent years, with particularly high levels witnessed in over the last two winters. It is no surprise to find that the system is at its most vulnerable at times of highest demand, but just how much of the risk is contained in a small number of extreme demand hours is worth emphasising.

The results also demonstrate why a healthy de-rated capacity margin is required in order to accommodate periods of higher than expected peak demand. These insights can be used when analysing the levels of forecast de-rated margin in the investment dynamics modelling. That is, given a particular level of forecast peak demand and de-rated capacity, it is worthwhile applying additional ‘stress tests’ when forecasting levels of de-rated margin to ensure that the system is robust given the uncertainty surrounding winter peak demand and availability of conventional generation.

3.7 Models applied to the GB market

A number of studies have been commissioned in recent years to investigate GB energy market performance and, in particular, answer questions about generation adequacy, costs and security of supply risk [23, 95–97]. This section contains a selection of key projects which have perhaps made the biggest impact in GB, that is either had a substantial influence on policy or have been the focus of significant press attention. The difficulty with reviewing work carried out by energy consultants is that, in most cases, model development is carried out in-house and typically details about methodology are not published. That said many of the results and associated sensitivity analyses can be examined and discussed.

3.7.1 OFGEM's Project Discovery

Much of the most recent work has focused on the capability of the current market framework to deliver a secure electricity mix whilst meeting firm emissions reduction targets and renewable energy obligations. A prominent example of this is the recent work by the GB regulator, OFGEM; their Project Discovery consultation [96] examines the *"prospects for secure and sustainable energy supplies over the next 10-15 years"*. This objective is explored using a scenario analysis with additional *"stress tests"* being applied. Four core scenarios are considered, all devised with decarbonisation as the ultimate goal (although this goal is not met in all scenarios), each with alternative outlooks for key factors such as economic and renewable growth, gas imports and commodity and carbon prices. An example of one such stress test is a situation when there is little or no wind output in a system with a significant wind penetration (up to 30 GW in the 2020 green scenarios for example). The outlook for security of supply risk levels[7] are assessed via de-rated capacity margins. The plots do not show past values of the de-rated margin so there is no way of knowing whether the situation is getting better or worse. The outlook is good for the near-term (3-4 years out), in contrast there is a significant erosion of margins after around 2013 followed by a slow recovery out to 2022.

A static approach to modelling is taken whereby an investment forecast is made assuming future gas and coal prices and then key indicators such as electricity prices are calculated. It could

[7]The report also considers security of supply risk for up-stream fuel supplies, namely gas imports from Russia, but this review examines their results concerning generation adequacy only.

be argued that a model of this type overlooks the market feedback mechanism that is present, whereby investors adapt their year-on-year investment decisions based on current market conditions; this includes projects in the pipeline, time delays, fuel prices and so on. Some of the other headline conclusions include that wholesale prices are expected to rise (although only average prices are shown and the volatility is not considered) with higher prices for consumers. These are to some extent offset by benefits of the low carbon electricity sector which may lead to lower bills in the long-run beyond the 15 year time horizon of the model. These additional costs come from a variety of sources such as fuel price increases, new capacity builds and network reinforcements, with the last two sources amounting to £200bn of cumulative investment in some scenarios.

3.7.2 Pöyry

It is widely accepted that the introduction of more variable output forms of generation such as wind will lead to higher wholesale price volatility which will impact on investment. One study which addresses these issues is that of Pöyry [98]. This study was commissioned in 2008 and with selected results published in 2009. Little information about the modelling methodology is in the public domain (e.g., [99]), although it is worth exploring the content of what is available owing to the influence the report had on government-informing bodies such as the Committee on Climate Change (CCC) [100].

The study takes 2000-08 wind and demand conditions, known as the *'mini-Monte Carlo'* iterations [99], and maps them onto 2020 and 2030 installed wind capacity in order to predict the contribution of wind to meeting demand and also assess the impact on market prices and thermal plant utilisation in both the GB and Irish markets. Price volatility is assessed via a price-duration curve (PDC) in both 2020 and 2030, the results for GB are shown in Fig. 3.12. The 2030 graphs show both negative and high wholesale prices. In times of high wind the price may become negative when there is a surplus of wind over demand or export capability in a given transmission-constrained area. On the other hand, at times of low wind the prices are expected to reach very high values as thermal peaking units will be seeking to recover their investment costs while operating at low utilisation, this is something which is also noted in OFGEM's Project Discovery. Mapping these changes to investment, the Pöyry report concludes

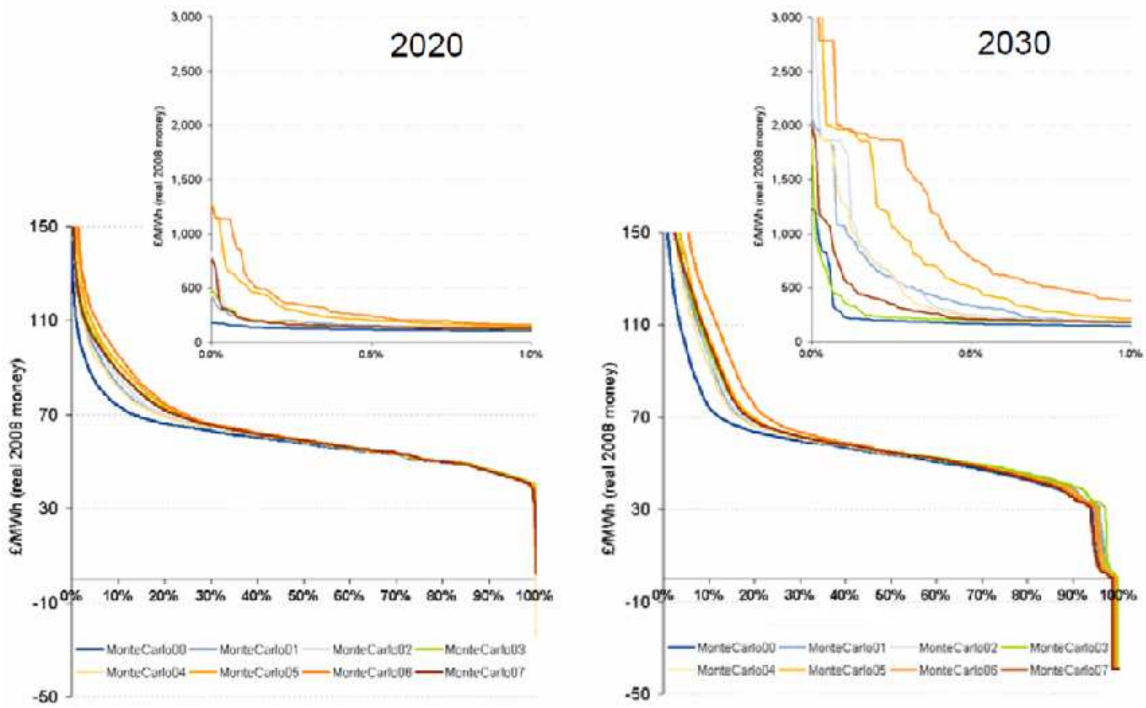


Figure 3.12: *Price-duration curve results for GB market in the Pöyry study [98] (used directly).*

that *"power stations built now will face a future of not only far lower load factors expected as a greater portion of the electricity but also dramatically increased uncertainty of revenues than at present"*. Interestingly the uncertainty of revenues is reduced in the Irish market. This is owing to a capacity payment mechanism, which forms part of the existing market design being included in the simulations.

The Pöyry model uses a mixed integer mathematical program which takes a number of inputs including historic demand; simulated wind output (36 locations in total across GB and Ireland), wave (using data from a Carbon Trust study [101] across 5 regions); simulated tidal barrage output (using profiles provided by DECC for the River Severn area); conventional plant availability (taken from half-hourly Maximum Export Limit data from the GB BM Reporting System [102]); realised commodity price data; and reserve and transmission zone data (7 GB zones modelled). The model time-step is hourly and plants are split by characteristics such as fuel, technology, efficiency, LCPD issues and capacity (although no derivations of plant cost curves or short-run marginal costs are provided). Using this data historic GB wholesale prices are simulated as a method of model validation. The simulated SMP provides a good fit during base hours, yet is below realised prices at peak (an example amount of £8/MWh is quoted). This is described as *"the value of capacity"* and it is attributed to peaking plants bidding in above their short-run marginal costs in order to recover fixed costs. As a result, simulated future prices are adapted to include a price *mark-up* in tight supply-demand hours. This amount is a multiple of the historic *mark-up* calibrated in order to allow peaking units to recover their fixed costs, however derivations of this multiplier are not provided explicitly. By including this price *mark-up* during scarcity periods, this allows for the expected wholesale price to climb above the SMP in periods of reduced capacity margins.

This method of price simulation has a significant impact on investment; investors in mid-merit and peaking units (i.e., CCGT and OCGT) are able to recover their fixed costs from the energy-only market because they can bid anything they like during scarcity periods, although a shortfall in revenue for very low-merit plant in reported. Base-load generation such as CCS coal and nuclear is treated as *"non-market determined"* and so growth in this type of capacity is made in addition to CCGT and OCGT investment in order to maintain an adequate capacity margin (although if periods of overcapacity occur, these plants will be closed if they are unable to recover their FCs). Thus in a model such as this, where the aim is to report the impact on prices of high penetrations of wind, security of supply concerns are less severe assuming the price and bidding strategies are permitted to operate in this manner.

3.7.3 Redpoint

The work carried out by Redpoint in reports such as [65, 95] and the many before it by the same group, is very relevant to the generation adequacy debate in GB because it has informed government departments such as the Department of Energy and Climate Change (DECC) and the Department for Business Enterprise & Regulatory Reform (BERR), and will thus influence policy. The most recent publication [65] has looked at options for GB market reform and others have included decarbonisation of the power sector [65], renewable support schemes [103] and the dynamics of generation investment [104]. All of these look at investment in generation capacity in some way and the results presented are typically very detailed. In [65] an outlook to

2030 with de-rated capacity margin, expected energy unserved and probability of brown out[8] is used to measure the level of security of supply risk. Additional policy mechanisms are modelled and their impact on reducing carbon emissions intensity and mitigating security of supply risks are measured. A wealth of sensitivities are also included. The headline result is that under current market arrangements de-rated margins will deteriorate after 2012 - with the situation becoming critical around 2018 - and remain below 10% for the remainder of the time horizon. Expected energy unserved will also increase dramatically after 2020, this is shown in Fig. 3.13. The large oscillation in unserved energy with frequency 1 year and amplitude 2-3 GWh after this date is curious as investment and retirements are reasonably linear throughout this period. It was argued by Redpoint [105] that the high values of unserved energy are on account of power shortages during summer when low wind resource periods coincide with large volumes of conventional generation being offline for maintenance.

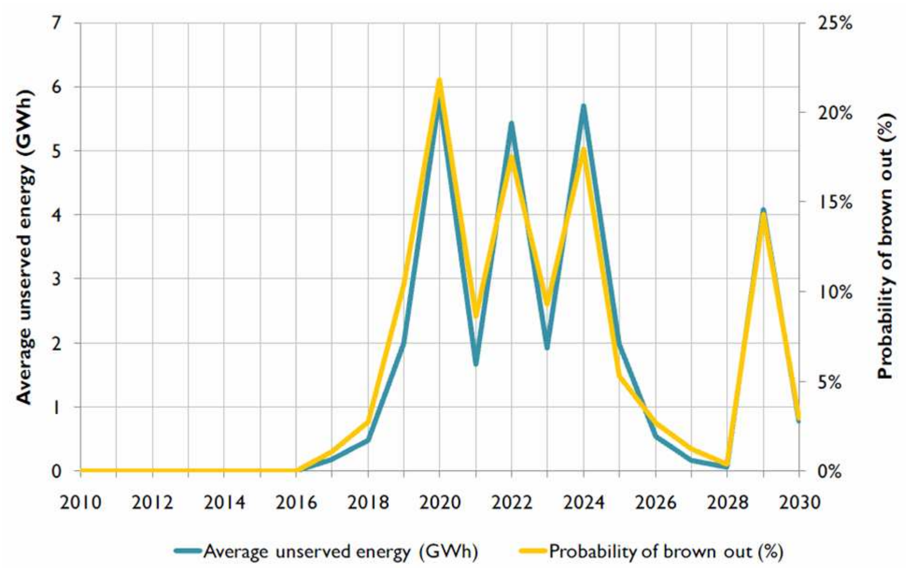


Figure 3.13: *Expected energy unserved results for GB market in the Redpoint study base case [65] (used directly).*

Policy schemes which successfully promote renewable generation lead to a reduction in wholesale prices and therefore insufficient conventional generation is built. To address this issue, a capacity payment is introduced from 2018 (the simulation year when significant risks arise) via a payment fund given by $FPD \cdot (1 + TDRM) \cdot (TFCO)$, where FPD is the forecast

[8] A brown out is defined there as *"a drop in voltage for some customers but without necessarily a full outage"* [65].

peak demand, $TDRM$ is the targeted de-rated capacity margin and $TFCO$ are the fixed and capital costs of a new OCGT plant. The pot is the same size despite the volume and type of capacity installed (annual payments will vary based on expected peak demand only). However the pot is spread wider when there is more capacity connected (and available), and therefore individual payments are less, and vice versa when less capacity is available. It also requires a target de-rated margin to be calculated. All generators receive the payment based on year-round availability and not at peak. It could be argued that a year-round calculation does not encourage generators to be available at peak (e.g., moving maintenance to summer), yet the payment does demonstrate that an additional revenue stream, which is less volatile and more predictable than the wholesale price, leads to more efficient investment in generation. Interestingly the additional revenue required to cover OCGT fixed costs via the capacity mechanism is found to be £7/MWh for an OCGT with amortised total FCs of 60/kW/yr and target de-rated margin of 10%. This figure is not too dissimilar to the required *mark-up* in [98]. Here a *mark-up* is added to the wholesale price simulations in the base case scenario (i.e., without capacity payments) and assuming this can be avoided in a market with capacity payments, increases to consumer bills as a result of the capacity mechanism being introduced are offset by prices being less volatile in periods of supply shortage.

In addition to the capacity payment, a targeted capacity tender is modelled. The tender period is annual and it is aimed at a *"small subset of generating plant or demand-side response"* [65]. This could be viewed as an extension to the existing annual reserve tender of 2 GW currently procured by the GB SO (cf. sub-section 3.2.2.3). Simulation results show that the tender serves its purpose and the target de-rated margin is achieved and the security of supply risk is *"significantly reduced"* [65]. The savings to customers are dependent on how the reserve tender market operates. More precisely, does the reserve capacity get called upon once all energy-only resources have been exhausted, or can an economic condition be specified over which the reserve volumes can be tapped in order to prevent prices reaching unacceptably high levels? In the former case, 'hockey-stick' style bidding (cf. sub-section 2.3.6) may still occur with an additional unnecessary cost to the consumer. Although not discussed in this report, a tender strike (or utilisation) price would prevent this from occurring, i.e., exercise the option to use tendered reserve, thus paying the strike price rather than the high wholesale price.

Details about the modelling methodology are hard to determine from reading the most recent publications alone, although assuming the same logic has been used throughout this series of work, some details can be determined by reading previous reports by the same firm. For instance, no detail about how the wholesale price mark-up function is calculated is given in [65], however in [104], the model *"adds an uplift factor which relates to capacity margins and has been calibrated against historic prices"*, which would suggest an approach similar to [98]. Other model features such as how the investment decision process is carried out is mainly provided through schematics describing information flows and software used with specific details omitted.

3.7.4 UKERC

Finally, a comprehensive review of the literature by the UK Energy Research Centre in [23] looks at the risks associated with investing in generation. The levelised costs of various *"leading"* forms of generation are discussed with sensitivities on key inputs such as discount rate, fuel price and capacity factor reported. A discussion about investor revenue and risks is included with points covered similar to those displayed in Table 2.1. Investment cycles in generation investment are seen as a key source of electricity price risk with timing of investment seen as critical. This last point is extended to look at the value (in terms of avoided revenue losses) investors can gain from waiting, in particular waiting for a policy change. This is interesting in light of the recent announcements regarding market reform [67]; if investors decide to perform these strategic actions, the market may see a investment hiatus until policy has been finalised.

3.8 Chapter summary

In this chapter a review of the GB ESI to date has been provided. This included a history of its industry structure in Section 3.1, with particular attention paid to the period post-liberalisation in Section 3.2. In Section 3.3 and 3.4 focused on the generation adequacy problem in GB and reviewed some past experiences. They key findings of this review were the evolution of market design from a pool to a primarily bilateral trade-based market, and the breaking up of large incumbent generating firms during this time. Another key finding was how the STOR market forms a vital part of current market arrangements in GB.

Next in Section 3.5 a historic generation adequacy risk calculation for GB was presented with results and discussion provided in Section 3.6. The results also demonstrated why a healthy de-rated capacity margin is required in order to accommodate periods of higher than expected peak demand, and that the scale of relative risk has changed in recent years, with particularly high levels witnessed in over the last two winters. Absolute values were also reported: the average LOLE over the 10 winter period was 0.06 hrs/yr with a standard deviation of 0.01 hrs/yr. A sensitivity analysis on the probability distribution for available conventional generation, showed an order of magnitude increase in these figures when the underlying distribution for available conventional capacity reduces.

Finally, in Section 3.7, an exploration of some of the key works which have looked at the GB market was provided, this included the work of the GB regulator OFGEM together with energy consultancies Pöyry and Redpoint. The key findings of these works were 1) more variable output forms of generation such as wind will lead to higher wholesale price volatility which will impact on investment and 2) that under current market arrangements de-rated margins will deteriorate after 2012 - with the situation becoming critical around 2018 - and remain below 10% out to 2030.

Chapter 4

Modelling Generation Capacity Investment in Electricity Markets

In the previous chapters an overview of various electricity market designs and generation technology characteristics was provided. This chapter presents an investigation into the many modelling techniques used in academia and industry. Particular attention is paid to studies which have addressed the issues of generation capacity investment and generation adequacy in liberalised energy markets. This includes a review of methodologies, the handling of uncertainty, market power and investor expectations and risk preferences. Finally, a discussion about the scope and value of using techniques from dynamic control theory to model the investment market is included. This will set the scene for the next chapter where the methodologies used in this work are presented.

4.1 System modelling

The transition to a privatised ESI together with targets to decarbonise electricity generation has given rise to many new and interesting challenges. By modelling these complex systems using the latest theoretical methods, understanding can be improved and better judgments made about what the future might hold. As described in [106], the most commonly used modelling techniques follow three main trends: optimisation models, equilibrium models and dynamic simulation models. These frameworks can be applied to a number of interesting problems within the electricity market, including the short-term unit commitment problem, market design, capacity expansion planning and congestion management. Plainly some models cannot be positioned within a single category because they have characteristics which make classification less straight-forward. However the representation forms an excellent basis for expanding the discussion.

4.1.1 Optimisation models

Optimisation models are the most common mathematical framework used in the energy sector today. This structure is applied in many situations owing to its flexibility, computational tractability and the numerous numerical solution methods available which can be applied to solve a range of difficult and complex problems. General research in the field of optimisation has led to significant improvements in the computational time needed to solve such problems.

These methods can be applied to formulate an objective function as a maximisation of expected profit for new generating capacity investment subject to constraints on a number of (uncertain) decision variables. This can be extended to modern portfolio theory to maximise profit subject to returns on the firm's existing portfolio of generating plant. Methods such as those devised by Bloom and Gallant [107] can be used; their primary goal was to obtain the optimal expected production of the generating units at each interval subject to matching the LDC and general linear constraints. Once the problem has been formulated, various optimisation techniques are available, including column generation methods (e.g., Danzig-Wolfe [108]) and active set methods [109], which solve smaller problems that expand (but never to an uncomputational size) to optimality. Applications of optimisation models go beyond the scope of generation adequacy to include optimal power flow, hydro scheduling and construction of bid and offer curves.

4.1.2 Equilibrium models

Equilibrium models are founded on the concepts discussed throughout Section 2.3. They are popular in electricity market modelling on account of their flexibility. For instance, both individual and aggregate supply functions can be modelled. They can be applied to a variety of market structures, with oligopoly generation markets a popular application (an approach originally developed in [10] and since applied by others, e.g., [110]).

There are two main kinds of equilibrium models that exist, namely *supply function equilibrium* (SFE) and *Cournot equilibrium*. In the former, market participants make strategic decisions on price and quantity whereas in the latter participants decide about quantities only and prices are determined via the inverse demand function. Both methodologies are based on computing the market equilibrium and can be applied in perfect or imperfect competitive markets.

In equilibrium models, under the perfectly competitive case, the overall level of capacity is typically determined by assuming that the scarcity rents received by peaking generators are just high enough to recover their investment and fixed operating costs [17]. Recall that scarcity rents for peaking generators occur during periods when total available supply is not sufficient to meet demand, and thus the price reaches the VOLL. The volume of each type of capacity (i.e., base-load, mid-merit, peaking) within the overall level of capacity is chosen so that that total profit for all generators is zero in the long-run. This can be represented mathematically using an optimisation with capacities as the decision variable and an objective to maximise gross consumers' surplus less the total costs of production, including both total fixed and variable operating costs, e.g., [111]. If the market is not perfectly competitive, for instance in an oligopoly, then the overall level of capacity is determined by profit maximisation. In this case, a firm may not increase capacity (the decision variable), even if long-run profit is above zero. More precisely, if the addition of new capacity triggers a reduction in scarcity rents across the firms existing fleet sufficient enough to incur fixed costs, then the firm will not invest. This can be represented mathematically using an optimisation similar to the perfectly competitive case. Here the objective is to maximise profit which is the price received for selling power less the cost incurred producing it (e.g., [111] uses a SFE model of this type).

Cournot equilibrium models are traditionally reasonably tractable, e.g., [76, 77, 110, 112]. Supporters of this approach suggest that a Cournot game is a fair reflection of what is actually going on in some markets and *"can support detailed cost modelling, and do not usually suffer from the multiple equilibria common with other modelling approaches"* [113]. This approach has been found to model the events of the California crisis of 2000/01 quite well, yet they tend not to give a good representation of how prices are set in electricity markets in general and their price predictions have generally been too high [76]. That said, they can still further understanding of competitive electricity markets. A recent example of a SFE model is [15] where the level of price mark-up in a market with a number of competing firms is investigated. This modelling framework has been used to highlight the potential for market power in highly concentrated markets. For example, in [78] the potential for submitting a supply function well above marginal cost in the highly concentrated E&W power pool was reported.

More recently these models have been used to assess the impacts of wind generation in competitive markets. In [17] the potential impact that a varying output profile from large volumes of wind generation (30 GW total; 19 GW onshore 11 GW offshore) has on half-hourly equilibrium prices is put forward. A market with six competing generating firms together with a duopoly of two symmetric firms is simulated. Results showed a significant increase in the variability of prices (and hence generator revenue uncertainty) for the high wind penetration. This was particularly so in peak demand hours, with the situation worse for a symmetric market. Although less common, SFE models that consider multiple firms can be adapted to model both aggregate system supply curves and also to assess more long-term market trends, perhaps embedded within a dynamic simulation model.

4.1.3 Dynamic models

The methodologies above can be embedded into a static or dynamic modelling framework. By choosing a dynamic model when considering the problem of generation adequacy, the outcomes of earlier stages of the simulation time horizon will directly influence the system dynamics at a later stage. For example in the case of power markets, current prices and their predictions are fed back to investors hence modifying the investment behaviour. The resulting investment decisions are then fed back to the pricing mechanism hence closing the information loop. This is precisely the approach taken for this project (Chapter 5). Consequently, a review of existing dynamic models applied to power markets has formed a vital part of this research. Dynamic simulation models are usually very computationally intensive and as a result some simplifications to reality must be introduced. On first inspection, this might be viewed as a disadvantage, yet by making reasonable simplifications control over input variables and model structure is achievable [114].

4.1.3.1 Introducing system dynamics modelling

One such area of modelling which has been widely applied in such situations is that of System Dynamics (SD). SD was developed during the 1960s by Forrester and colleagues at MIT where classical control theory was applied to various industrial systems exhibiting interacting growth and decline characteristics (or *"multi-loop nonlinear feedback systems"*) [115]. These included a number of public policy problems such as energy, health, environment, social welfare and security [116].

Although SD had been applied to electric power systems expansion before privatisation (see review in [117]), it really took off post-liberalisation and has been applied to the long-term generation adequacy problem to great effect. The work of Ford [117] in the US, and Bunn et al. [34, 118] in the UK are the main protagonists. SD continues to be a popular choice for electricity market modelling [119–123]. For example in [119], SD modelling is employed to address questions of stability and dynamics and draw attention to the idea that capacity expansion and wholesale prices are governed by the negative feedback connections between supply and demand.

4.1.3.2 SD to highlight investment cycles

In [124] a simulation of the western US power markets is presented. The model is used to assess construction cycles in generation capacity and the effect of long lead times and capital intensity under the prevailing Californian market rules. The study highlights the tendency for the timing of private investment to be late with resulting dramatic price spikes and reliability problems. There is then a subsequent boom in capacity construction, followed by a bust cycle. SD models are described as a good method for illustrating these cycles because there is a need to understand the feedback mechanisms present in such a system [124]. Broadly speaking, the SD work of Ford draws the same conclusion about market-driven investment in new generating capacity; fixed capacity payments are needed to provide the extra revenues to generators.

Expanding further on the commodity cycle problem, it is widely accepted in economics that oscillations occur as a result of long-term dynamics in supply and demand or some form of system shock (e.g., a reduction in demand due to economic depression). In the case of an SD model of capacity expansion, these are either endogenous (e.g., demand) or exogenous shocks (capacity expansion or contraction) [115]. Such random variations may occur for a number of reasons, such as lumpy investments and individual strategic moves [125].

In [126] a SD framework is used to examine the over-investment effect of a pricing mechanism. The test cases are ‘energy-only’ with and without a price cap and a capacity payment mechanism. The model is quite stylised, with a SFE pricing mechanism assumed. Investments take 5 years to build and the model includes an option for investment to be abandoned during the 4^{th} year of construction. An interesting conclusion of the paper is that 'energy-only' markets do provide adequate investment signals under *"ideal conditions"*, presumably meaning in a market which is in long-run equilibrium, and capacity payments lead to over-investment and oscillations in capacity, although supply shortages occur less frequently.

4.1.3.3 Incorporating market structure

SD models can also be formulated to compare how a system comprising of a number of technology types, or *vintages*, impact on investment dynamics. Vintage modelling is a method used in macroeconomics to differentiate between types of capital (here, generating plant) each with their own rate of productivity (here, thermal efficiency). For example, long-term electricity market dynamics is the focus of [121] where a simplified capacity mix has been modelled using a number of vintages. The capacity within each vintage is time-dependent and is described by the rate at which new capacity enters the vintage (i.e., new builds) and the rate at which capacity leaves it (i.e., retirements).

Assumptions about market equilibrium have also been tested in SD models, with [127] describing a feedback model of generating capacity investment decisions. Results from simulations that assume an equilibrium versus a nonequilibrium liberalised market are compared. They show that assuming an equilibrium market can lead to strategic errors in the investment decision process and significant differences in realised wholesale prices and reserve margin. The main difference between the assumptions underpinning each model relates to the aggregate investment rate. The equilibrium investor assumes that the aggregate industry investment will evolve in order to maintain a target reserve margin. Here an investment is deemed profitable if the expected wholesale price is greater than the total levelised investment cost. As expected prices are modelled as a function of reserve margin (prices increase as the reserve margin decreases), the feedback between investment and prices is captured.

In [128], two neighbouring perfectly competitive electricity markets are modelled. The system consists of four thermal generating technology types in each region, with an interconnector linking the two. The relationships are defined in terms of a causal loop diagram, with an endogenous generation investment decision component. To reflect the interaction between spot markets, two feedback mechanisms are present. The first is a negative feedback loop between installed generation capacity and expected spot prices, with construction lead times included. The second, also a negative feedback loop, defines the relationship between the two regional markets, with regional demand impacted by power trading. When imports increase, the residual demand in the importing region reduces, which in turn reduces electricity spot prices. Three types of market design are modelled and the impact on investment, system reliability and price dynamics is assessed. Capacity margins and traditional power system reliability metrics are used to assess 30 year generation adequacy risk in each region. An 'energy-only' market is used as the base case design in each region. This is compared with markets with price caps, and a capacity obligation with a forward ICAP market and price cap. A mixture of symmetrical and asymmetrical markets designs is also trialled. Results show that an asymmetric market design can lead to some unintended consequences in a neighbouring region, particularly if the market designs differ between energy-only and price cap, where capacity transfers to the non-price cap region in times of supply shortage. It is stated that *"similarities of approaches lead to better performance in terms of price level and variability and in terms of reliability"* [128]. Of those trials where an energy-only market is adjacent to an ICAP market, two things are noted. Firstly, the ICAP market does not rely on the energy-only market for back-up, and secondly, the energy-only market does not gain an advantage from being interconnected with a market that has a formal adequacy policy in place.

4.1.3.4 Incorporating investor profiles

In [122] a stochastic model to analyse long-term characteristics of electricity markets and to *"determine the fundamental factors contributing to (long-term) supply adequacy"* is presented. The focus is on how future prices inform investment, and argues that *"futures provide market participants not only an instrument to hedge price risk, but are also an indicator of future supply and demand conditions, which are critical in making an investment decision"*. The wholesale market price is modelled as a function of reserve margin. Demand is modelled by a seasonal factor, in one case this is constant and in another it has monthly variation, with a stochastic element to represent uncertainty about future load. Various time delays are tested with two investor models, namely forward- and backward-looking investors with and without an ICAP market. The market performs more efficiently with a forward-looking investor in all cases (lower prices and less capacity oscillations).

In [129] an SD model incorporating credit risk theory is used to differentiate between firms and concepts from game theory are called upon to include strategic behaviour from market participants. The model is applied to a simplified model of the Spanish system (mixed hydro-thermal, no wind) comprising two large companies, one medium, three small and several IPPs (as one pseudo-company). A gaming element is included by assuming strategic actions under a spot market design. A fully competitive market is also tested along with a market with large volumes of forward contracting. Interestingly out of the four unit types available to investors, (nuclear, coal, CCGT, GT) only the latter two are chosen. This may be because the cost characteristics are unfavourable for nuclear and coal (no cost information is shown so it is difficult to establish whether these are suboptimal choices of investment for all utilisations). As with the actual Spanish system, the market arrangements in all cases are such that a capacity payment is included in addition to energy market revenues. In fact it is noted that without this mechanism, many of the GT investments would not take place.

In [130] a dynamic model incorporating game theory with multiple firms participating in the market is presented. This work is applied to real electricity markets, in total eight countries are considered. Three scenarios for levels of competition are included, namely a fully competitive market and a market consisting of strategic firms. More precisely, the strategic scenarios are two-fold; one takes production levels of rival firms into account and the other models a *Stackelberg game* (where the leader firm moves first and then the follower firms move sequentially). As a general point, if the market is perfectly competitive, firms' decisions about quantities will follow a *Nash equilibrium*, where each agents strategy is the best one in light of the other agents strategies. An assumption of the Nash equilibrium is that agents have perfect information about the strategies of others. Note that perfect competition is not necessary to obtain a Nash Equilibrium, in fact it is actually a more powerful tool in imperfectly competitive markets. The focus is on the impact that investment trends have on prices and the environment in the long-term (2000-2050). The investment decision element of the model is dynamic with investment based on feedback obtained from the most recent market information. Constraints on production capacity, electricity trade and GHG emissions are also taken into account.

4.1.3.5 Information flows

The advantage of SD models is that the flows of information between elements of the market can be captured. Fig. 4.1 shows some of the key information flows in power markets. In this particular depiction there is a high penetration of wind present; the diagram shows how the availability of the wind resource influences price dynamics which feeds back into the investment decision and hence impacts on the economics of conventional generation. Information on capacity in development also flows though to current investment decisions as do current and future market prices. The volatility of the wind output and prices impacts plant utilisation. This can lead to a change in available capacity on account of plant mothballing or premature retirement. System demand is inherently linked to wind output as both are heavily dependent on the weather. Surrounding the main diagram are other factors, such as technology costs and investor risk preferences. These play a crucial role, and although not explicitly labelled with feedback arrows (to avoid over-complicating the diagram), some of these factors will alter as a result of power market conditions. For instance, economic growth could be harmed if wholesale prices rise too high (perhaps in response to increasing fuel and emissions costs) or, alternatively, investor cost of capital grows as a consequence of increased revenue uncertainty. The key point here is that a great deal of information flows through the market and changes in one area will propagate through the system, sometimes instantaneously and sometimes with delay. These dynamics are of great interest in SD modelling.

4.1.3.6 Model validation

An important stage of any model aimed at simulating a real system is validation. Owing to the nature of liberalised energy markets, where commercial sensitivities typically mean a limited volume of data is available to call upon, this stage can be difficult. However many models have managed to perform some form of validation, which have enhanced the model results.

[131] produces 8 scenarios for the future development of the electricity system. Validation against the past is included whereby the model is run against 1998-2006 German market data and found to match well against both investment and prices. It is argued that a dynamic element of the model enables identification of time delays and market imperfections, which can temporarily lead to the market being out of equilibrium [131]. A classic causal loop diagram

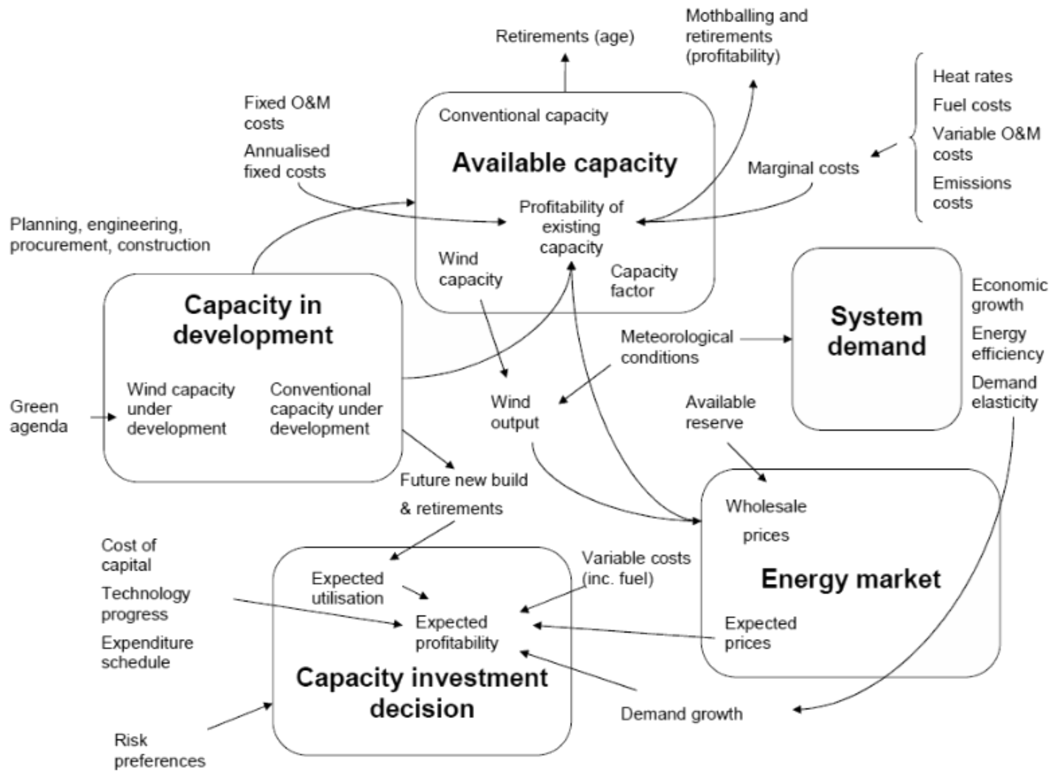


Figure 4.1: *Information flows in the generating capacity investment market. Inspired by [119].*

is included, which encapsulates the feedback and information flows similarly to in Fig. 4.1. Unlike the representation in Fig. 4.1, causal loop diagrams also highlight (using a double line crossing an arrow) system time delays. An arrow between two variables, say, from a to b, indicates that a relationship exists and a positive (or negative) sign at the end of each arrow implies that a small positive change in a subsequently leads to a positive (or negative) change in b [128]. A good example of this is seen in [121] and a version of this is shown in Fig. 4.2. Delay d_1 occurs before the investment decision due to the irreversibility of investment requiring sufficient certainty about expected returns on investment. Delay d_2 occurs after the investment decision due to construction lead time. The circle arrow in the centre of Fig. 4.1 containing a negative sign indicates negative feedback implying system balancing [128]. To avoid over complicating the diagram, much of the additional detail in Fig. 4.1 (including the impact of wind generation) is omitted.

Another noteworthy example where an investment model has been applied to an existing system is presented in [110]. A stochastic dynamic model of a liberalised energy market using a

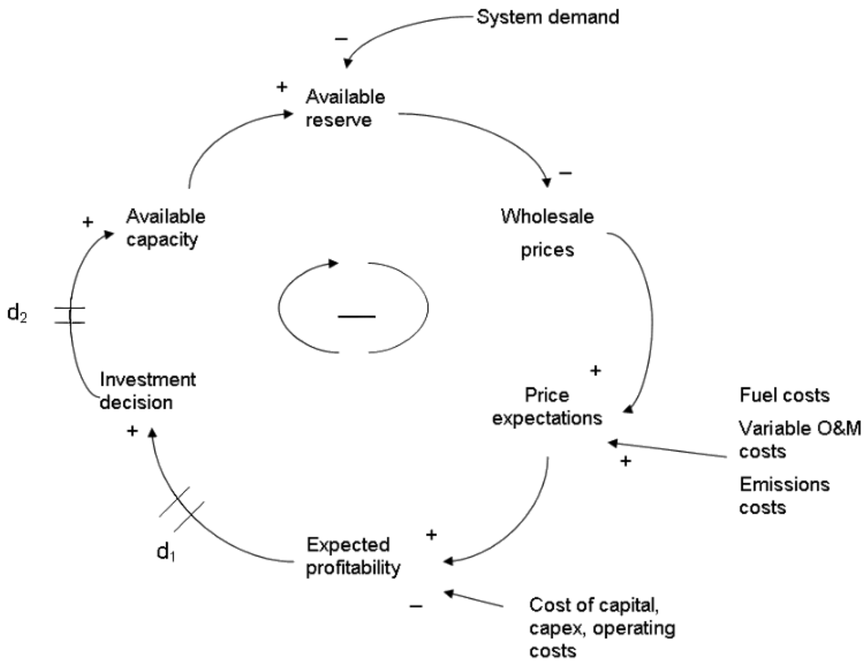


Figure 4.2: *Causal loop diagram for the generating capacity investment market after [121].*

Cournot model is applied to the Finnish system. The goal is to replicate investment and production trends over the period following liberalisation in 1995 by focusing on the years 1996-2006, and also to test the impact of varying key input parameters and model logic. To reflect the oligopolistic market structure over this period, three competing firms are modelled. These firms must decide on production and investment in each year of the simulation time horizon. The authors describe the market as *"segmented"*, with load partitioned into base and peak periods and firms make their decisions based on these. The motivation for this is based on empirical evidence of *"interdependency between prices (or demands) in the base- and peak-load segments"*. As a result, two inverse demand functions, which are calibrated to the Finnish system using available data on *"prices, quantities, and price elasticities"* are considered when computing Cournot equilibrium. A sensitivity where market segments have *"cross-price elasticities"* is included. Here base- and peak-load inverse demand functions contain a positive cross-price elasticity coefficient applied to quantity in the opposite market segment. Interestingly this has little impact on investment trends. Other sensitivities included varying the depreciation rate of capacity, price elasticities and investor planning horizon. A planning horizon of 6-7 years provided the best match with realised investment timings and volumes. The model predicted more base-load investment than in reality but overall volumes and timings were reasonably well approximated. Furthermore, lengthening the planning horizon from the base case of 5 years, led to *"more ambitious and earlier investments"*. This could be due to simulated prices and production needs further out being higher than in the near-term on account of positive demand growth raising production requirements and prices (capacity depreciation rates are held at zero for this test); thus sending an earlier signal for investment. Varying demand elasticity had only a minor impact, whereas using a positive rate of capacity depreciation led to a delay in investment and production was shifted toward peaking generators. The latter occurs as an increase in capacity depreciation *"increases the opportunity cost of the investment due to shorter life expectancy of production capacity"*. Results also showed that modelling production cost as a function of total installed capacity (i.e., marginal cost of production varies in the economic dispatch), increases the performance of the model in predicting overall investments and production costs. Interestingly, all types of capacity considered were modelled with a one year construction delay. This may explain why base-load capacity (e.g., hydro and nuclear), which have longer lead times than peaking capacity, was a more popular choice for firms than in reality.

Other dynamic models which have been applied to real systems include [112, 120, 130]. In [112] a dynamic simulation model was used to calculate the profitability of new plants. It addressed plant lead time and included a capacity payment calculated using the same method as under the E&W Pool. It concluded that a market with this capacity payment leads to capacity oscillations and highly volatile wholesale prices. In [120] a SD model is applied to the NordPool market where there is a spot, futures and real-time market.

4.1.3.7 Answering policy questions

More recently, SD models have been applied to inform the debate on decarbonisation in the electricity sector. [132] focuses on investment dynamics in a system with carbon markets and a view to reduce the emissions in the US Western Interconnection. Another example is given in [133] where a Tradable Green Certificate (TGC) mechanism is tested and the primary focus of the model application is on price dynamics rather than investment. It is noted that *"only after some time does the system respond with enough new capacity to allow the price to decline to values that one would expect from a fundamental analysis"*. Moreover, the time lags of construction can create highly volatile oscillations in the TGC price during the period when installed capacity is increasing.

The work of Peña [134] uses a SD model to look at regulatory policies to address generation adequacy and environmental targets. Here two capacity mechanisms are tested, 1) capacity payments with a price equal to the average fixed cost not covered by selling energy to the energy market (25% of costs assumed to get covered in market), and 2) capacity market with the sloped demand curve design described in [54]. The elastic demand curve is calibrated such that the capacity market price matches the capacity payment price when the target reserve level is met. In addition to capacity mechanisms, the effect of renewable subsidies are tested. Over the 25 year time horizon of the model, the capacity market leads to a more stable reserve margin, better efficiency and more peaking capacity investment.

4.1.4 Where this thesis fits in

A variety of modelling approaches have been discussed throughout this section along with a number of applications. It it therefore worthwhile briefly pausing to reflect on where the work presented in this thesis fits into the current literature and contributes to the field of knowledge.

Firstly, this thesis looks to build on existing dynamic investment models by applying the techniques covered in sub-section 4.1.3 to create an investment market model for the GB energy market. Furthermore, taking inspiration from models that have been applied to existing markets such as Ford [135], Bunn [112], and others discussed earlier, this thesis can extend the field of knowledge of applied system modelling. In addition, it addresses the issue of high penetrations of wind power in such models and how this impacts on investment dynamics, which remains an active, but not yet fully understood, area of research. For instance, this work will compliment studies such as [17] where the impact on market prices and investor revenues of a high wind system is considered. Finally, like the models discussed in sub-section 4.1.3.7 (and in the case of GB, Section 3.7) this work can inform policy makers interested in the response of energy markets to policies promoting wind generation.

4.2 Uncertainty modelling

"There is nothing more certain and unchanging than uncertainty and change." - JFK.

The future is uncertain, we can be sure about that. System modellers must make judgments about what the future might hold in order to inform the debate about the best course of action for today. In the case of electricity market modelling a number of techniques and philosophies have been developed. In the following sections some of the methods influencing this work will be discussed.

4.2.1 Monte Carlo method

In situations where there is a high degree of uncertainty about the model inputs, then the Monte Carlo (MC) method is a common approach. With modern computing power, repeated random sampling within a model of a large and complex system can be achieved. This sampling provides a method to reduce model error. Further, the larger the sample, the closer the sample average will be to the expected value (law of large numbers), therefore a high number of MC simulations is recommended. Examples of studies which have incorporated the MC method include [54] where 25 simulations of a 100 year time horizon are run because the model randomly samples economic growth and weather. Also Pöyry [98] used this method to simulate future GB wholesale market prices based on random sampling of wind speeds and other uncertain variables.

4.2.2 Stochastic processes

A *stochastic process* is a set of random variables ordered according to some index, typically time. Using stochastic processes within a model is useful when there is belief about how a variable will evolve over time but it is in part random and unpredictable. The evolution of such a process may also be referred to as a *random walk*. The field of stochastic processes is vast, and a detailed discussion of the mathematics underpinning it is beyond the scope of this thesis. However the interested reader could consult Chapter 2 of [32] for a comprehensive introduction to the subject.

The type of stochastic process of interest here is the *Wiener process* (or *Brownian motion*), which provides a basis for modelling a broad range of stochastic processes. These are continuous-time Markov processes (where only the current value is required to forecast the next value of the process), and have independent increments. Many of these properties can be visualised by inspection if the generalised formula for a one-dimensional Brownian motion or Itô process:

$$dX_t = m(X_t)dt + v(X_t)dW_t, \quad X_0 = x > 0, \tag{4.1}$$

where X is the state process, W is a one-dimensional standard Wiener process, m and v are given deterministic functions. In plain terms, the function m is the expectation (or drift) and v is the volatility (or "noise"). By definition $W_s - W_t \sim N(0, s-t)$, where $s-t$ is the length of the time-step. That is, the expectation of the volatility component is zero. Or put another way, the expected change, dX_t, is precisely the result of $m(X_t)$. Furthermore, if the functions m and v are unchanging overtime (i.e., the process has constant statistical properties), this is then termed a *stationary process*.

Examples of the use of stochastic processes in electricity generation investment modelling abound. This is because it underpins real options analysis and can be employed to forecast uncertain parameters such as future demand and wholesale market prices (e.g., [136]).

4.2.3 Real options analysis

Although not employed in this thesis, real options analysis is a popular approach to modelling uncertainty in generation investment. Therefore it is worthwhile discussing a number of key examples from the literature. The basic principle of real options is that the value of investment can be maximised by choosing the optimal time to invest, i.e., there can be value in waiting. Perhaps the best-known text on this field is that of [32] where the mathematical background along with a firm's decision process in the face of uncertainty are discussed.

In [137] a dynamic model incorporating real options theory is presented. The underlying value of a project is modelled as a stochastic process. It is argued that models of this type are developed to represent the whole life of an investment: of initiation, running and abandonment, not just the net present value (NPV) [137]. The objective of the investment decision can be to maximise social welfare (centralised decision) or maximise profit (generating firm decision). The paper discusses the introduction of price feedback in the decentralised case and finds that a price cap below the VOLL leads to under-investment (or in the context of real options, project delay). Capacity payments lead to earlier investments but also over-investment in peaking capacity.

In the follow-up work described in [138] a stochastic dynamic model is presented. The model is stochastic owing to uncertainties such as load growth. Wholesale prices are modelled as stochastic processes with fuel prices deterministic. These dynamic and stochastic characteristics are the key motivations for choosing real options over a static NPV approach. The benefit of not having to determine a risk-adjusted discount rate is mentioned, although an estimated risk-adjusted discount rate is required for the simulations.

Investment in CCGT plant with either flexible or constant generation profile is considered (the study example is the Nordic market) and the influence of a market with and without capacity payments is analysed. The model works backwards from the last year of the planning horizon (10 years out) with the decision in each year being the optimal one, given the stochastic inputs of demand and available generation. The decision to invest is taken annually and is limited to a single CCGT plant over the planning horizon (the purpose of this analysis is to establish the effect of market structure on investment thresholds, not on total levels of investments). The agent is assumed to be a price taker and the influence of a new investment on an existing portfolio is not considered. As the wholesale price is a function of installed capacity and demand, they take the logical step to constrain off the influence of the price process beyond the end of the time-horizon (i.e., assume that at some point before then, new capacity will enter to lower prices in the long-run).

Two types of capacity payment were considered: fixed payments or a simple function of peak load and available capacity. In the second case, the payment only occurs if the capacity margin falls below 7%. Capacity payments increase expected profits for both the disapatchable and non-dispatchable technology options. Furthermore, optimal investment thresholds are reduced (i.e., investments will be triggered earlier). Interestingly, investors wait for the variable capacity payment to grow until they trigger investments and therefore a higher payment is required in order to enduce more timely investment.

Incorporating uncertainty into systems modelling is a challenge faced by all disciplines. For power market modellers, the methods discussed above provide a means of addressing uncertainty in a simulation environment. Stochastic modelling is a well-established methodology which bestows both confidence when applied in a simulation environment and a degree of realism when interpreting the results. It allows foresight, but like any system with a high degree of uncertainty, these predictions must be communicated and interpreted accordingly.

4.3 Scenario modelling

An alternative to tackling uncertainty using the simulation methods outlined above is to take a scenario approach to modelling. This approach was originally established in the early 1970s by Royal Dutch/Shell as part of development and testing of its strategic planning [139]. These types of models can incorporate multiple pathways by using scenarios to analyse alternative futures. Traditionally, this approach seems to be favoured by policy makers because, owing to the largely deterministic and hence uncomplicated modelling framework, they are able to visualise the different futures over a range of input and policy scenarios (e.g., [96]).

When building dynamic models, a typical approach taken is to incorporate a number of scenarios pertaining to “core” inputs with other inputs either modelled deterministically or stochastically. For example in [65] a number of policy scenarios are analysed in a dynamic GB electricity investment market model and their impacts on carbon intensity reduction, security of supply and consumer welfare are assessed. In [21] medium and long-term levelised costs projections for a range of generating technologies available in GB are analysed against a number of scenarios for key cost drivers. More precisely, the common style of low, medium and high cost scenarios for commodity prices, build timings, technical data, capital costs and technological maturity is employed.

A similar method to scenario analysis is sensitivity analysis, and broadly speaking the objectives of both are analogous. A sensitivity analysis is usually used to test the robustness of a model under a range of values for its input variables. In contrast scenarios are typically constructed prior to implementation and consist of only a small number of discrete values for the inputs being varied (e.g., with low, medium and high fuel price or demand growth scenarios commonplace).

4.4 Agent modelling

Market environments facilitate competition between participants (i.e., generating firms). Depending on the number of effective sellers in the market (monopoly, duopoly or oligopoly for example) the actions of individual firms play a key role in market outcomes. In modelling terms, these firms are commonly referred to as 'agents'. Each agent will have their own attitude toward risk and expectations about the future state of the system. When modelling a liberalised energy market environment and the investments therein, the aggregated decisions of the agents directly influence the evolution of the system. The methodology chosen to model these interactions and decision processes relies heavily on the model scope; for example whether a single agent or the interaction between multiple agents is being investigated.

For example in [135] three types of investor are modelled within an SD model of a generation investment in California, namely *believers*, *pre counters* and *followers*. Believers only include plant under construction once it comes online (i.e., they do not forecast its arrival earlier in the simulation when it is under construction). *Pre counters* assume all plant under construction will complete and include it in the simulation as soon as construction commences. *Followers* have a *"herd mentality"*; they will not begin building new plant, even if a it is deemed profitable, until others have begun construction elsewhere. The results of the interactions between these investor types is quite interesting. Under a scenario with just *believers*, a boom-and-bust investment cycle is witnessed, this can be attributed to investors continuing to initiate construction when there is adequate capacity in the pipeline to meet demand growth. This behaviour leads to prices being significantly damped owing to a large capacity margin when the excess new construction comes online. A simulation with solely *pre counters* leads to investment that just keeps pace with demand growth and the pattern of prices is more oscillatory; high prices in peak periods when supply and demand margins are tight and lower in low demand periods when there is excess capacity available. In a mixture of all three cases, a construction boom occurs early in the simulation involving all three investors. The *pre counters* withdraw early in the boom (owing to their forecast of new build), with *believers* and *followers* continuing to invest. As a result, the installed capacity surpasses the level necessary to meet demand growth. In fact, out of all the simulations only the purely *pre counters* run does not display boom and bust investment characteristics.

Other SD models which implement a multiple agent investment market include [129] where credit risk theory is used to differentiate between agents; more precisely an NPV is calculated for each investor based on *"a different endogenously-calculated discount rate"*. [127] presents a SD model with a duopoly investment policy based on a desired market share together with IPPs who invest based on *"managerial optimism"* about market share and expected investment by rivals.

Another notable multi-agent model is presented in [118]. The model pays particular attention to the trading of plant between agents, i.e., purchase or dispose. The market is deemed in equilibrium when no plant trading occurs. The model is constructed based on a Cournot game, whereby agents decide on quantities to bid based on what they expect other agents to submit. The second stage is then a Bertrand game which is analogous to a balancing mechanism and day-ahead market; here agents set prices and the demand-side decides the quantities at that price. A number of "markets" are simulated, namely base, shoulder and peak, to which each agent (there are 24 in total) must determine (via a profit maximisation optimisation) how much capacity they will bid into each market.

The studies cited above are designed to model multiple agents. However there are many models which model a single agent under the assumption that the behaviour of this agent is a good representation of the market, and investment decisions are modelled as aggregated. Many SD models take this approach including [119, 121–123, 128]. The issue of modelling aggregated investor behaviour where, in reality, there are many agents competing with one another and each agent has his own expectations about the future must be addressed. The next section describes two different approaches which do this.

4.4.1 Expectations hypothesis

In addition to the attitude of market participants toward uncertainty and the risk associated with it, the manner in which participants gather and use information to form their expectations about future market conditions is an important consideration. In particular, for a model where price signals drive investment in capacity, expectations about future price and its distribution must be considered [121]. This phenomenon was investigated in [125] where an interactive computer simulation model of a simple electricity market, with users acting as generating firms undertaking strategic investment decisions, was used to determine the rationality of capacity investment decisions. As highlighted in [20], results showed there was a tendency for users to 1. initiate investments during high price periods, and 2. ignore capacity under construction and construction time lags, thus leading to periods of surplus capacity and low prices.

A key element of commodity cycle theory is the expectations of participating agents. A number of expectation hypotheses have been developed, and perhaps the most hotly debated is that of *rational expectations*. The development of this hypothesis removed the possibility of cycles described by the cobweb theorem [20]. Followers of rational expectations believe that it may not be possible to design stabilising policies to influence individual agent behaviour, which when aggregated, dampen cycles and achieve market equilibrium. Further, the hypothesis argues that economic models of commodity markets do not assume enough rationality [140] and therefore cannot be used as a policy tool. The *natural rate hypothesis* [141] states that there is a fixed relationship between the aggregated supply and the difference between the expected and the actual market prices. This implies that in order to change the level of installed capacity in relation to natural levels, policy must influence the difference between actual and expected prices [141]. For rational expectations to apply, a number of critical factors must be present. These are not discussed here, but details are provided in [141]. Connected to this is the argument that central predictions about market outcomes (e.g., [67, 96]) affect agents' forecasts and therefore influence the course of events, as summarised in [142]:

"*The problem of invalidation of a public prediction arises because the public prediction may affect the agents' expectations and thus become a determinant of their behavior. Under these conditions the public prediction becomes itself one of the factors which determine the future course of events: because a public prediction has been made, the event which occurs at the specified time is different from the one which has been predicted and which would actually have occurred if the prediction had remained private.*"

Therefore results from any economic forecast model entering the public domain must bear these implications in mind.

Plainly boom and bust type dynamics will be less severe if investors have rational expectations. For instance in [135], under the exclusively *pre counters* situation investment keeps pace with demand growth because investors take into account new builds in the pipeline. In general they require an intimate understanding of the market and consequently take time to form. This is arguably not possible in most electricity markets owing to their relative infancy and ongoing modification [54].

In light of this, *bounded expectations* is perhaps a more appropriate model to consider; this theory assumes that agents are rational, but only within the limits of the information available to them. Because their information may be bounded they can sometimes take decisions that appear to be irrational according to traditional theories [143]. Bounded rationality is used in [114], where two SD models are presented, one with an annual generation capacity investment decision and the other with a capacity utilisation decision (i.e., how much capacity to release to the market). The models *"rely on simplifications that differ from reality"*, although they highlight the cyclical nature of prices and the SD formulation allows program users (actually a class of final year undergraduate Engineering and Economics students) to make active investment decisions during the simulation. They propose that users use a heuristic investment decision process, which is bounded and based on experience. The *"decision function"* assumes agents *"use a feedback strategy to adjust their capacity towards a desired capacity."*

Other contrasting theories to rational expectations includes *adaptive expectations*. This theory assumes that agents form expectations using past trends and any errors in their earlier predictions will continue. In [121], it is argued that *"adaptive expectation formulations based on exponential smoothing of past values and trend extrapolation can well replicate aggregate forecasting behaviour."*. Here a similar argument to [54] is used to discard the rational expectations hypothesis *"based on the fact that firms do not know the exact specification of the system, and therefore, the structure of forecasting models and many parameters must often be judgmentally estimated"*.

4.4.2 Risk aversion

Investors attitudes to risk play a large role in behaviour and investment outcomes. A risk averse agent would rather receive a given amount of wealth with certainty than the same amount of wealth, or perhaps higher, on average but with variance around this quantity [54]. In this section some of the techniques used to model risk preferences within decision processes are discussed.

4.4.2.1 Utility functions

Risk aversion can be described by a *utility function* which maps wealth to utility (or satisfaction). Investors who have a diminishing marginal utility are termed risk averse; if investors have a linear marginal utility then this is termed risk neutral (Fig. 4.3). The most common form used to represent risk averse investment is an exponential curve owing to its downward sloping (concave) property [54]:

$$U(x) = a - be^{-Rx}, \tag{4.2}$$

where R is the risk tolerance (the larger R, the more risk averse), x is the expected monetary value, and a and b are calibrated so zero results if expected wealth is zero and utility above a certain wealth is 1 (i.e., full satisfaction). So for example, consider a situation, s, where you can win $£10$ with probability 0.5 and win nothing with probability 0.5. The expected outcome is $E(s) = 0.5 \cdot £10 + 0.5 \cdot 0 = £5$, however a risk averse participant may opt to take $£x$ $(< £5)$ with certainty. This x is the *certainty equivalent*. In (4.2), if $x = 3$, then $R = 0.18$ can be obtained by solving $U(3) = 0.5$, $U(0) = 0$ and $U(10) = 1$, and the utility function's concave form is achieved. Under a risk averse utility, the expectation of the utility for a given expected wealth is less than the utility of the expected wealth. This is the reverse inequality of Jensen's inequality applied to a convex function. Jensen's inequality states [144]: if f is a convex function over the interval I and $x_1, x_2, \ldots, x_n$ are in I, then

$$f\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right) \leq \frac{f(x_1) + f(x_2) + \ldots + f(x_n)}{n}. \tag{4.3}$$

An example exponential utility function is shown in Fig. 4.3. In this case, the function has been calibrated by forcing $U(5) = 0.7$, $U(0) = 0$ and $U(10) = 1$."*indicating somewhat but not extreme risk aversion*" [54]. The two black dots show an example of the function obeying Jensen's inequality.

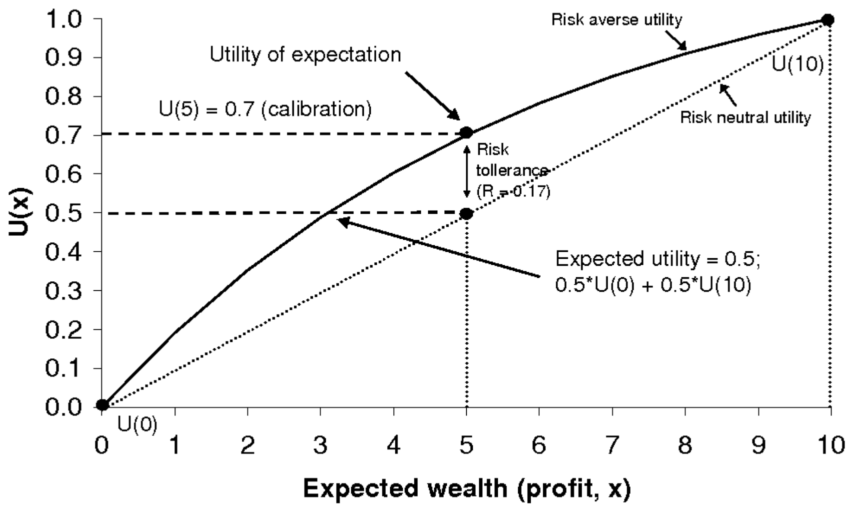


Figure 4.3: *Example of expected utility function. Function is calibrated so* $U(5) = 0.7$ *(in (4.2)). This yields* $a = b = 1.225$ *and* $R = 0.17$*. Inspired by version presented in [54].*

4.4.2.2 Prospect theory and risk seeking behaviour

Prospect theory [145] is an alternative to expected utility criteria. It argues that the positive utility function does not tell the whole story, and there is actually a negative utility function to consider. When investors stand to lose money they take bigger risks in the hope of recovering their losses. In the context of power markets, this could be the difference between a firm mothballing a plant as soon as it starts losing money, or keeping it available in the hope that prices will improve. This is an important consideration in an investment market model that considers price feedback as the main driver of boom and bust cycles. If this type of behaviour is present, it will depend heavily on the financial stability of the firm; wealthy generating firms can afford to take risks but new entrants and small players cannot. The middle ground contains participants who are concerned with the reference point, or point of inflection, which is where the utility function changes from positive to negative.

An example of a utility function with a reference point at zero is shown in Fig. 4.4; investors are risk-seeking when expected wealth is below the point of inflection and risk averse above it. By moving the reference point above zero the more likely the participant is to be risk-seeking. This behaviour is perhaps best summarised in [146]: "*the higher the reference point the more likely the agent is to be a risk-taker (think of the notorious cases of Barings' Nick Leeson and*

Société Générale's Jérôme Kerviel), and hence the greedier she is.".

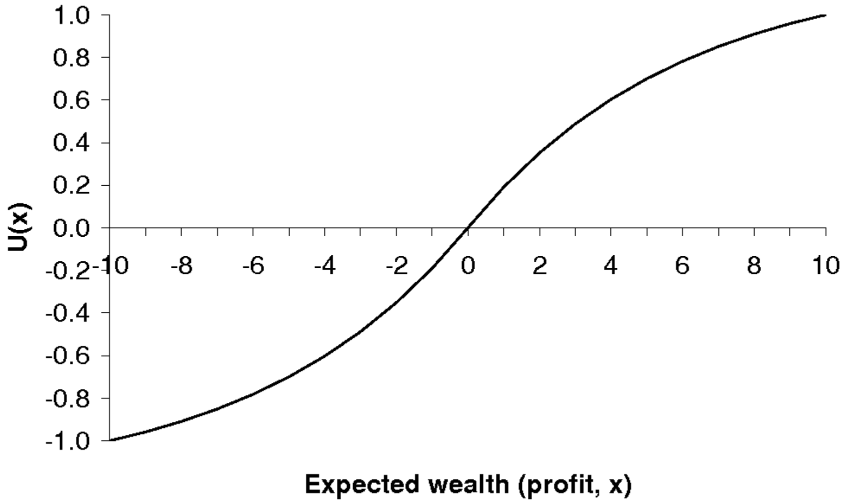


Figure 4.4: *Example of utility function with reference point at zero. For expected utilities greater than zero the graph is as in Fig. 4.3. For negative wealth the function is calibrated so $U(-5) = -0.7$ in (4.2), giving $a = b = -1.225$ and $R = -0.17$ (i.e, an exact reflection of the positive utility function in the x and y axis).*

4.4.2.3 VaR criteria

Value at Risk (VaR) is a common criterion used in finance when investors are concerned with the extremity of realised returns below expected values (i.e., they are risk averse) [147]. Generally speaking, given a distribution for project value V (the random variable), and defining the level of risk aversion by q, the VaR is defined as the value v_q such that $p(V \leq v_q) = 1 - q$, i.e., v_q is the value of the q^{th} percentile. An example of this concept is shown in Fig. 4.5.

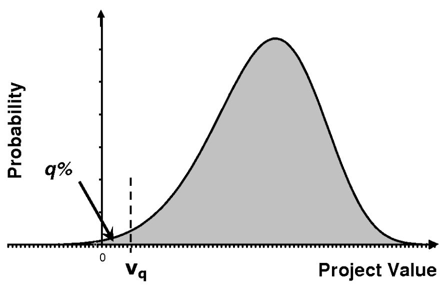


Figure 4.5: *Example of VaR decision criterion acting on a distribution of expected project values.*

Despite their popularity, VaR models do have a limitation relative to other risk criteria (such as Min Max Regret) as VaR analysis must consider probabilities of outcomes, but typically these probabilities are not well known. An alternative approach would be to use Min Max Regret whereby an investment decision for a particular realisation of the system is analysed. Regret is defined as the difference between the net benefit incurred when an investment decision is made for a realised state of nature, and the maximum net benefit across all feasible strategies that can be obtained under that realised state. For instance, investing in nuclear generation for a realised carbon price that is lower than expected might have a higher regret than investing in gas-fired generation under the same price. However, the Min Max Regret criterion only considers the worst regret for each alternative, which makes its results very sensitive to the exact set of scenarios that are considered.

The Conditional Value at Risk (CVaR) test is more sensitive to the shape of the distribution for V. It is said to provide a more coherent measure of risk than VaR as VaR is incoherent as regards to the addition of risks [147]. If q is the CVaR level, then the expected shortfall at the q level is the expected value in the worst $q \cdot 100\%$ of cases. So for high q the most profitable but unlikely values are ignored, but for low q only the worst cases are considered. In the latter case, an investment is made if the expected project value is positive conditional on the revenue being below v_q, with q the confidence level, and investment is triggered if $E\left[V|V < v_q\right] > 0$.

4.5 Modelling market power

The opportunity for firms to abuse their position in the market and perform strategic or anti-competitive actions in order to increase profits must be avoided (cf. sub-section 2.3.6). A number of models have looked at this element of power markets (e.g., [77, 78, 118, 148–150]) with the aim of assessing market power in current markets (e.g., [78, 148, 149]) or looking at methods to mitigate against it (e.g., [3, 13]). Market power can be assessed in a number of ways. Perhaps the most standardised method used is by inspection of the industry Herfindahl-Hirschman Index (HHI). HHI is the market concentration calculated as the sum of squares of individual firms' share [148]:

$$HHI = \sum_{i=1}^{N} (s_i)^2, \tag{4.4}$$

where N is the number of firms and s_i is the market share of firm i. Depending on whether the proportion s_i is expressed as a decimal or whole number, the HHI can range from $1/N$

to 1 or 100 to 10,000. Typical guidelines suggest that a HHI above 0.18 indicates a highly concentrated market and above 0.1 indicates a concentrated market (e.g., Office of Fair Trading merger guidelines [151]). The inverse of the HHI gives the effective number of competing firms. A rough calculation for the total installed capacity in GB market, which in 2011 consisted of around 44 firms, shows a HHI of 0.095, with inverse 10.5, indicating a concentrated (or oligopoly) market. Comparatively, the HHI for Sweden and Singapore are reported as 0.32 (8 firms) and 0.27 respectively [148]. Moreover, a cross-country comparison study for Europe for 2003-05 [152] concluded that the GB market compared favourably with other European countries such as Germany and Spain. The report states that *"the market structure in Great Britain can be seen to be the only one largely conducive to competitive outcomes"* [152]. For instance, the average HHI values based on available installed capacity were 0.107, 0.191 and 0.279, for GB, Germany and Spain, respectively.

The classical definition of market share, s_i, is *"output of supplier divided by total market output"*. Therefore the HHI calculated above is not the classical definition because it has been determined using total installed capacity, which is different from total output, yet it is popular with regulators as it is easy to calculate. In order to determine the true HHI index, one would have to calculate the total output (demand met) at each market settlement period (or average over all periods) and determine each supplier's contribution. A method to calculate this Dynamic HHI (DHHI), is described in [153]. This paper also provides a measure of *"gaming incentives"* in a uniform spot market. The incentive depends on the shape of the supply curve, the market share of participating generators and demand elasticity. The formula calculates the benefit from withdrawing plant; namely the increase in profit for the firms remaining plants from price *mark-up* plus the fixed cost savings as a result of plant withdrawal.

Some other market power detection methods are listed below. This is by no means a complete review of market power indicators and the interested reader should consult [13] for a review of the subject:

- The *Lerner index* (LI) measures a firm's ability to exercise market power by calculating the level of price *mark-up* they can induce. If the price set by generator x is P, then the LI is given by $LI = (P - SRMC_x)/P$ where $SRMC_x$ is the generator's true SRMC. Note that the *price-cost margin index* (PCMI) is a similar metric,

$PCMI = (P - SRMC)/SRMC$. If the market is perfectly competitive, then both metrics will be zero for the generator producing the last unit of production needed to meet demand. However the use of LI or PCMI is problematic given the difficulty in determining generator marginal costs [13].

- The *supply margin assessment* (SMA) is a type of *pivotal supplier indicator* (PSI). PSIs, which are 0/1 binary variables, attempt to incorporate both supply and demand conditions to assess whether a particular generator is 'pivotal' in serving demand [13]. The SMA for generator x is made at peak demand and is given by $SMA_x = (D - \sum_{i=1|i\neq x}^{N} C_i$, which is demand ($D$) net supply from all other generators' capacity, C_i, in the market. If this value is positive then generator x is a pivotal supplier and $PSI = 1$, otherwise $PSI = 0$. The SMA has been criticised for a number of reasons, one of which is that it is calculated for a single demand hour only [13].

- The *residual supply index* (RSI) is like a PSI but operates on a continuous scale and is more expressive of actual market conditions. It is defined as the ratio of total ICAP less the generator's relevant capacity, RC_x, over demand (D): $RSI = (ICAP - RC_x)/D$ (as a percentage). Note that D is the sum of full load plus purchased ancillary services and RC_x is the generator's capacity minus it's contracted obligations. This value can be calculated for all load hours and provides a more complete assessment than SMA. A value below 100% indicates a pivotal supplier, and a value greater than or equal 100%, indicates a non-pivotal supplier. [13] states *"that on average an RSI of about 120% will result in a market price outcome close to the competitive market bench-mark"*. Here the "bench-mark" is the estimate of the market price in a perfectly competitive market.

- The *concentration ratio* (CR) is a measure of total output produced by a given number of firms relative to the size of the market. For example, the CR capacity ownership in the GB market for the 'big six' generators is around 70%, which again indicates an oligopoly market.

When modelling market power, or perhaps more precisely, detecting market power in existing markets, monitoring can occur at one of three periods. That is, either in advance, close to real-time or retrospectively. The choice depends largely on the metric being used and hence data requirements. For instance, those that require data on realised prices (e.g., price trends or congestion analysis) can only be carried out *ex-post*, whereas those looking purely at market structure (e.g., HHI) are largely dependent on generating capacity data and can be carried out *ex-ante*. Supplier type indices (e.g., LI) can be carried out during all three time periods although the degree of realism will vary because sometimes data estimation is required. Furthermore, the degree of monitoring ease depends on market structure. For instance, markets where the majority of trades occur bilaterally will be more difficult to monitor than those where the majority of trading occurs in a pool or exchange (although problems remain if this data is also not in the public domain). An in depth discussion on this topic is beyond the scope of this thesis; [13] provides an excellent discussion on the need and design for effective market monitoring.

4.6 Timing and lumpiness of investment

A large volume of work has been carried out in the economic literature investigating the relationship that time lags (e.g., [154]) and lumpy investment (e.g., [155]) have on commodity and business cycles. The main cause of time delays are construction time or planning permission hold-ups. Further, as discussed in Section 2.6, in a market situation they can be a result of strategic actions by investors such as delaying investment until more information becomes available and revenue uncertainty is reduced. There are many theories explaining why investors tend to delay; on the one hand delays are costly because future revenues are discounted and *ceteris paribus* investors would like to receive profits today. On the other, by making an investment, there is a risk it will fail and thus the opportunity for future profit is lost and so there is an incentive to delay [156].

In the case of power systems, lumpiness of capacity new build and retirement is unavoidable - generating units traditionally come in sizes of the tens to hundreds of MWs range. In addition when the cost of new capacity is weakly concave then economies of scale will encourage large-scale investments which are few and far between [157]. Some assessments have reported that underlying engineering procurement and construction costs are also cyclical [21], thus exacerbating the issue.

Strategic delays tend to be modelled (in the economic literature) in a real options setting, where the primary focus is the option *value* of waiting (sub-section 4.2.3). Early work focused on instantaneous investment timing alone [32] but more recent literature has addressed the impact of time delays [156–160]. In [156] a simple dynamic model based on real options theory shows that build delays can amplify *"the natural cyclical nature of the economy"*. The work presented in [158] examines the timing and intensity of investment in order to *"shed some light on the dynamics behaviour of investment"*. The model presented builds upon the existing theory of [32] by addressing two decisions for the firm; when to invest and how much to invest, with alternative lumpy or incremental investment models presented. It is argued that by *"adding the choice of investment intensity enriches the model"*. An interesting conclusion is that investment under uncertainty not only delays investment but also lowers intensity when investment is incremental. By comparison this is not necessarily the case for lumpy investment. This type of work is relevant here because generation adequacy is not only reliant on timely investment, but also adequate investment. An important distinction between lumpy and incremental investment is made; *"that new capacity is not compatible with old capacity"*. More precisely, incremental capacity is installed as an extension to existing capacity whereas lumpy investment is either a replacement for retired capacity or separate from existing capacity. Under this definition, there is scope for both incremental (e.g., wind farm extensions) and lumpy (e.g., new CCGT plant) investment in electricity generating capacity.

This work is extended in [157] to a situation where a firm is seeking to devise an optimal investment strategy in the face of uncertain demand growth. The formulations presented are more representative of a dynamical system where equations defining installed capacity, stochastic demand and time-lagged new builds form a stochastic differential equation (SDE) of *excess* capacity. This can be analytically solved in order to find a closed-form optimal investment strategy for a given set of input parameters. More precisely the investor objective is to minimise capacity costs, which include maintaining current capacity and installing new capacity, with demand being modelled as a stochastic process and the price of capacity constant. Interestingly, retirements to existing capacity are not considered and are therefore not included in the SDE. The price for capacity is constant regardless of system excess capacity (which can be negative) and therefore a *"trigger"* excess capacity level determines project initialisation rather than the usual trigger price. It is precisely the assumption of constant price that enables a closed-form optimal control investment strategy to be derived. In electricity generation investment, the price for capacity is not constant, although the work does present some interesting results which help explain why time delays, investment lumps, uncertainty and the interactions between them can lead to suboptimal timing of investment. The authors seek to address a number of issues. Firstly, does uncertainty delay investment? Secondly, does greater uncertainty raise the trigger price for investment and reduce the trigger price for abandonment? And finally, does uncertainty *"depress"* investment activity? The two examples presented show some surprising results. In the first test case, where the time delay for new build is set at one year, the excess capacity trigger level decreases (investors wait longer) with demand uncertainty and also increase the investment lump size. In the second example, where delays are set at eight years, the trigger excess capacity level increases (i.e., investors move earlier) and in fact *"decreases the firms sensitivity to uncertainty"*. This rather unexpected result that higher demand volatility leads to accelerated investment is explained by the firm's *"desire to contain the risk of capacity shortage"*. This is less easy to control with longer time delays.

[161] extends [160] to look at mitigating cycles in a endogenous growth model with time delays. The structure of the model presented (a social planner problem similar to [160]) indicates that some of the mathematical assumptions about the utility function described in [160] can be relaxed and a *"closed-loop policy"* (CLP) developed. A CLP, like an open loop policy, is an optimal policy to induce an desirable pathway for the control variable. However CLP uses the feedback mechanisms present to gain a deeper understanding of the *"economic implication of these models"*. Assumptions concerning utility function are relaxed compared with [160], however the structure is still quite stylised - for instance, the production function is linear and the utility function is homogeneous. The damping effect produced by the CLP is that those investors who foresee future capacity coming online (rational expectation with perfect foresight) and adding to supply will smooth oscillations.

Owing to the stylised nature of many of the models discussed above, it is difficult to use the ideas directly in a generation investment market setting. That said, the concepts presented can form a basis for formulating a dynamic model of the system, particularly the delay equation formulations presented in [162] and [157]. These works will be referred to again in Chapter 5, when the dynamics of the investment model implemented in this thesis are formally defined.

4.7 Applications of control theory

As the approach presented here is concerned with applying techniques from optimal control theory to model power market investment dynamics, it is worthwhile exploring if and how this methodology has been applied in power systems economics modelling in the past.

Feedback control theory has been applied to power systems in the past, within the more recent, but largely equivalent discipline of SD. To date, studies have addressed more short-term issues concerning generation resource availability (e.g., [163–168]). In contrast nothing in the literature has used classical stability analysis in order to design an investment market controller with the aim of mitigating investment cycles in the power market. This was the primary motivation for employing techniques from dynamic control in this thesis. That said, the concept of implementing a control mechanism in an SD model is not new. For instance, in [169] a simple control mechanism is implemented which uses a target level of installed capacity as the reference signal. If the market is not expected to achieve this target level, thus leading to *”extra required capacity”*, then an *“incentive signal”* is sent to generators in order to induce investment. The extra required capacity is estimated by a central body at the planning timescale far enough in advance to allow for construction delay (here time to build CCGTs). The incentive signal is an amount of revenue paid to generators separate to market revenues (i.e., capacity payment), designed to cover the expected proportion of investment costs of the extra required capacity not provided by the wholesale market. This additional revenue can vary year to year depending on the amount of capacity required and expected revenues received from selling energy on the wholesale market.

In [164] a control system model of an energy market with differential equations governing supply and demand is presented. Suppliers can choose to expand production when the system price is above production cost; equally demand can increase consumption when the system price is below the marginal benefit to consumers. Both forms of expansion are not simultaneous, which is reflected by the inclusion of a time delay constant in the differential equations. The state variables are the quantities and so the imbalance between supply and demand forms a third equation. In the base case, supply must equal demand at all times and in the second case the supply-demand equation can be temporarily unbalanced which will be reflected in the system price. Owing to the transparent nature of the differential equations (i.e., the production cost and marginal consumer benefit functions are known), the system can be analysed for its stability characteristics using classical eigen analysis, with the system deemed stable if the eigenvalues lie in the left hand side of the complex plane. In fact, the system modelled is a linear first-order so the eigenvalues (which are a function of time delays and linear coefficients of the cost functions) have real parts only and therefore the system is stable if these are negative.

Both of the control models presented in [165] and [166] are concerned with generator bidding strategies in spot markets. Using a Cournot oligopolistic market model, [166] provides a number of strategic bidding strategies for producers with alternative market expectations. In [165] a bidding system based on price and demand feedback is constructed and a optimal controller is designed with the goal of modifying the SMP in order to induce the necessary bidding behaviour to meet system demand. This approach of modelling the electricity market as a closed loop dynamic system leads to a deeper understanding of market dynamics and stability characteristics. Furthermore the use of optimal control theory enables us to develop strategies which can guide the system toward stability, which is not possible through the use of static models [170].

In [163] an SD model is presented where demand is modelled as a stochastic process and the simulated market price is a function of demand and available reserve. An optimal control problem is formulated with the goal of maximising the welfare (profits) for both the supply- and demand-side, with system reserve being the control variable. Interestingly, although the optimal control problem is designed to find the competitive equilibrium between supply and demand, the generator ramp constraints, or *"supply-side friction"*, led to prices fluctuating between the maximum (or *"choke-up"*) price and zero with no convergence toward the SMP. This high price volatility is seen as a fundamental part of competitive power markets. The supply-side friction is similar to the problem of lumpiness of new build and time lags during capacity construction; new capacity capable of meeting demand does not become available immediately and so there is a wait, *"which allows the firm to accrue a larger than the static competitive equilibrium profit for a while"* [163].

4.8 Summary of influence of literature

Drawing on existing knowledge has formed a valuable part of this research. For instance, the work in the field of system dynamics has heavily influenced project scoping and model implementation. The processes used in system modelling when defining the information flows and system time delays has provided a concise method of defining this particular dynamical system. This is important for communicating the model framework in a manner that can be understood by researchers across a variety of disciplines.

The concepts discussed in Section 4.2 can be used to address many of the key uncertainties that arise in power market modelling, particularly those that attempt to predict the future and influence policy. As will become clear later on, this has informed the approach to modelling uncertainties such as fuel prices, carbon prices, demand growth and construction lead times using well-established stochastic modelling techniques.

A review of agent-based models has highlighted some of the limitations of what can be learned from a single “representative” agent models, but at the same time it has demonstrated that there is value in exploring this case, and if performed in the right way, robust conclusions can be made. As is a common theme running throughout the literature, the principle of Occam’s razor should be applied from the outset; elements such a multi-agent logic can perhaps be added at a later date to increase the model’s scope. Furthermore, Section 4.5 on market power has increased understanding of how market power is monitored in power markets, and more appropriately, replicated in a simulation environment. Many of the market power detection methods are not used within the model, although indicators such as RSI and SMA give confidence that linking the degree of price mark-up to available capacity margin is a credible approach.

The concepts of risk aversion have influenced the modelling of investor decisions. By reviewing all available methods of representing attitudes toward risk, the most appropriate methodology can be applied. Furthermore, taking the perceived value of investment and translating that into an aggregated market response decision has been heavily informed by the topics covered here.

Timing and lumpiness form a crucial part of the investment cycle phenomenon and gaining an understanding of this is vital in any model of long-term investment trends. The models discussed have aided the formulations of the system (cf. sub-section 5.1.5) and have helped communicate the mathematics that underpin this model.

Finally, a review of existing applications of control theory has provided inspiration for the project, and assurance that there is value in taking this approach and that contribution to knowledge is achievable.

Chapter 5

Implementation of a GB Generation Investment Market Model

In this chapter the methodologies used to address the research question posed in Chapter 1 are presented. To begin, the fundamental model design concepts together with their application to a dynamic investment model of the GB power market are presented. Included are details about all elements of the first iteration of the model. The purpose of this iteration was to build on the preliminary work carried out by Häni [123] and apply the model to the GB investment market. Also included are the results and discussion accompanying this stage of the work. The model's ability to simulate the market trends witnessed in Britain since early 2001 are scruntinised. This acted as a basis for extending the methodology to account for high penetrations of wind power, which is the topic of Chapter 6.

5.1 Application of dynamic control to GB market

5.1.1 Basic concept

Under certain assumptions, fundamental economics dictates that market forces will match supply with demand in the most cost-effective way.[1] As discussed in Chapter 2, 'energy-only' markets rely on price signals, in particular price spikes and increased scarcity rents, to provide feedback to investors and incentivise the building of new capacity. To understand the effect these characteristics have on market prices it is useful to consider plant investment as a negative feedback control mechanism with market conditions, which can be used to forecast energy prices, acting as a feedback signal. This concept is depicted in Fig 5.1. Because the model is dynamic, current prices and their predictions are fed back to the investment block modifying

[1]These assumptions relate to the market being perfectly competitive. This includes: 1) the market has many participants, none of whom have market power; 2) demand is responsive; 3) there is perfect information flow; 4) participants make rational decisions; and 5) there are no systematic externalities.

the investment behaviour. The resulting investment decisions are then fed back to the pricing mechanism, hence closing the loop.

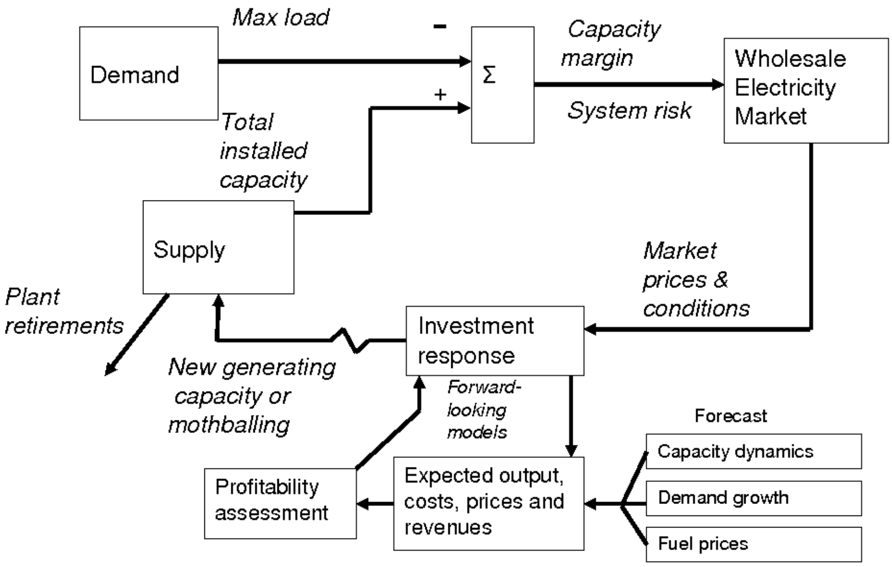


Figure 5.1: *Electricity investment as a control problem; investment can be viewed as a negative feedback control mechanism with current and future energy prices (as a function of generation capacity margin) acting as a feedback signal.*

The elements of a classical control problem can be observed in Fig. 5.1: they are the reference signal (capacity margin), the system gain (investment or retirements) and the feedback signal (market conditions, i.e., current and future prices). Further, the combination of inherent investment time lags, delays and uncertainty about the level of capacity coming online in the future can lead to generation capacity oscillations. It is well known in control engineering that time lags and uncertainty worsen system stability and create undamped oscillations.

5.1.2 Following Häni [123]

The Diploma project of Häni [123] was carried out at the University Of Edinburgh in 2005 and acted as a starting point for this research. It analyses how wholesale market prices impact on the decisions made by investors to install new or disconnect (mothball/retire) existing capacity. The electricity market is modelled as a dynamical system and the work by Visudhiphan [122] was the foundation. The main approach is to consider the capacity margin as a proxy for security of supply risk and use different available market instruments to attempt to implement a model that achieves long-term capacity margin stability (i.e., dampen oscillations).

The model is summarised by Fig. 5.2. An aggregated approach is taken where individual generators and consumers are merged into a single supply and demand component. The supply side is dynamic and is described by the following steady-state discrete system:

$$\left\| \begin{array}{rcl} x(m+1) & = & x(m)+u(m) \\ y(m+1) & = & y(m)+u(m) \end{array} \right\| \tag{5.1}$$

where $x(m)$ is the installed capacity at month m, $y(m)$ is the sum of the installed capacity and exogenous inputs plus new generation, both of which are represented by $u(m)$. By representing supply dynamics using a set of parallel difference equations, the implementation follows classical control theory. This allows the system state (x) and output, (y), to be measured at each time-step (m). Initial values of installed capacity for nuclear, coal, CCGT and OCGT are included, meaning a separate dynamic block for each technology. These technology types are then intrinsically connected in order to simulate the aggregate electricity investment market. Nuclear and coal are considered base-load capacity and CCGT and OCGT are considered peaking. Each technology experiences a time delay on account of the construction stage. The exogenous input in $u(m)$ is modelled by a random variable reflecting levels of uncertainty in supply and project abandonment. The new generation element is endogenous and represents new capacity added to the system (i.e., as a result of the investment decision process) or disconnected. The sign on this element is unrestricted: positive for plant addition, or negative for plant mothballing or retirement.

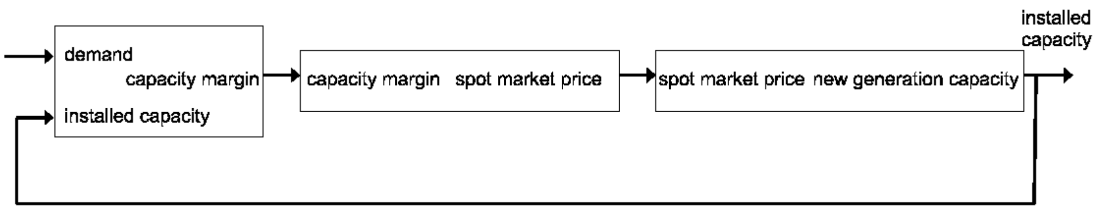


Figure 5.2: *Summary of Häni model; an electricity market with capacity and demand as the only variables influencing energy prices [123].*

Häni's spot market price formulation *"does not reflect influences other than pure market rules"* [123] and the energy price is a function of only two parameters: total demand and total available capacity. The price model also reflects capacity shortages or excesses by producing an infinitely high price when demand exceeds capacity and very low price when the capacity margin is large. The author states that *"a realistic model has been constructed; it will not produce* *"nice" results, but rather credible results for the input data given"*. This is an important point and is discussed further later. Simulations start with reasonable parameters; for example, a 20% capacity margin along with a random combination of peaking and base-load plants, each with different life expectancies. No assumptions about total system costs are stated explicitly, although the fact that plant annual capacity factors are computed using screening curves suggests an initial mix and amount of capacity consistent with the minimum cost of supply.

Only maximum monthly demand is considered in the simulation, which is reasonable as this is when the capacity margin will lessen and potentially induce price spikes. These prices may lead to a systematic over-estimation of revenues by investors if used as the investment signal. Fluctuations during a given month are well known and the detailed load pattern is periodic and seasonal. The scale of these fluctuations is similar throughout the year, thus enabling estimation of the actual values of base and peak demand (via a simple, linearly decreasing, LDC). There is a pre-defined pattern describing seasonal changes modelled using a region-typical demand curve for one year. The simulated monthly demand also contains two random variables. The first reflects uncertainties (e.g., weather and demand forecast error) and is sampled from a Normal distribution with mean 0 and variance specified by a fixed volatility factor. The second is a random variable which increases over time to reflect higher uncertainties for distant future periods. Annual demand growth is included, with a constant rate of increase assumed. Demand elasticity is considered to be virtually zero, although the final demand model includes elasticity and a demand cap to simulate disconnection or *"self-rationing"*.

The model uses the Internal Rate of Return (IRR) method when calculating investment decisions, i.e., the value of the discount rate which yields a NPV of expected revenue of zero. Information on total demand, total operational capacity, generator life expectancies and expected commissioning dates for capacity under construction are considered transparent and observable. More precisely, the IRR for each technology type is calculated prior to model execution for the full range of possible spot market prices. Then a minimal acceptable IRR is used as a trigger for investment. In fact, the market spot price is the only dynamic parameter in the model and hence determines the investment decision. Depending on investor behaviour, expected revenues can be calculated can be either *"current, past and maybe future market prices"* [123]. In the case of mothballing (and de-mothballing), a threshold market price, equal to the variable operating cost of the technology is used as the trigger price. Results showed that the outputs of the model (i.e., installed capacity in Fig. 5.2) are very sensitive to the parameters of the pricing mechanism, which is unsurprising given that revenues are inherently linked to this parameter. The impact of investor behaviour is also tested in the Häni model. Taking inspiration from Ford [135], *'believers'* and *'pre counters'* were implemented, and an *'average investor'* was also constructed, which was a combination of the two. As with the work by Ford [135], the more rational the investors expectations of new capacity coming online, the less overshoot dynamics were witnessed.

The purpose of the Häni work was not to replicate a specific market, but to act as a proof of concept for a dynamic control approach to modelling capacity margins in market environments. The relative success of implementation in the Matlab/Simulink programming environment was encouraging, although the model is quite stylised and in order to be applied to a real market environment, further development is required. Furthermore, some problems were encountered when implementing (5.1). These mainly involved the transfer of state information to other functions and blocks within the model (a crucial element of any state-space model). As a result, many of the aims, particularly those relating to the implementation of a control signal (i.e., capacity mechanism) could not be fully investigated.

5.1.3 The GB investment market model

Having gained an understanding of the work of Häni, attention now turns to this model implementation. It was decided that applying the concepts first considered by Häni to a particular energy market would prove extremely valuable and contribute significantly to this field of knowledge. The GB investment market was chosen for the application. In the remainder of this chapter, the first iteration of the model framework is presented together with preliminary results achieved from a hindcast simulation against historic GB market trends. This part of the work was carried out during the first 18 months of the research and allowed deeper understanding of the complexities faced when modelling a real system in a simulation environment, and the degree of realism that can be obtained from the results of such models.

When simulating a real system using a dynamic control model, one must remember exactly that; that it is not the real system but only a substitute for it. Using computer-based simulation models enables these complex systems to be studied in an academic environment, though the outputs of such models must be interpreted accordingly. They are usually based on a number of (typically simplifying) assumptions in order to keep the model tractable. Classical control theory recommends taking three steps when analysing a system:

1. Model the behaviour of the real system.
2. Identify problems such as instability (e.g., boom and bust investment cycles).
3. Optimise the controller to meet performance specifications (e.g., reliability/risk criteria/target capacity margin).

To understand and control these systems a mathematical model is required. As the systems under consideration are dynamic in nature, the input/output relationships are described by differential (or difference) equations (e.g., (5.1)).

5.1.4 Data considerations

When using results from a simulation model to make a prediction about the investment and generation adequacy dynamics under a particular market framework, it is helpful if the validity of the model can be tested, particularly if the outcomes of these models will directly influence policy decisions. In order to simulate past trends, historic market data must be obtained. The availability and quality of such data will determine firstly, the amount of validation that can be carried out and secondly the level of confidence that can be taken when comparing with history. Below is a list of the key data requirements for such a task together with resources:

- **Market price data** for the period before the introduction of NETA has proved difficult to obtain, therefore the validation work is constrained to a comparison with market trends post NETA (i.e., 2001 onwards). Historic Market Index Price (MIP) data is available via the GB price reporter ELEXON [171]; this a reflection of the short term wholesale price in GB and matches Reference Price Data obtained from the APX Power Group [172]. More detailed wholesale market price data is available via subscription to the power exchanges, but the prices charged are beyond the financial limits of a postgraduate studentship. Further, industry regulator, OFGEM, can be called upon to provide pricing data in its market liquidity reports (e.g., [8]).

- Databases such as the Economic and Social Data Service [90] can be called upon to provide historic data for a number of countries, of particular interest are figures on total installed capacity, generation by fuel type and peak demand. This data can also be obtained from NG's Seven Year Statements [27], which includes information on plant **capacities, commissioning year, retirements and mothballing**. Further, information on plant mothballing in response to market conditions is available from OFGEM [173]. As GB SO, National Grid (NG) publishes a Winter Outlook [66] in which it makes an assessment of both the security of both gas and electricity supplies.

- **Generating unit cost data** can be estimated using these and other published studies such as the macro-level MARKAL model [174] and, more recently, the extensive investigation into generation costs published by Mott MacDonald [21]. In fact, the latter has been used as input data for much of the recent modelling work commissioned by DECC (e.g., [65]). These resources report technological characteristics such as generic thermal efficiencies and annual plant availability.

- Details of **utility financial structures**, such as cost of capital and gearing ratios can to some extent be derived from company annual reports (e.g., [41–46]) or independent firms that have carried out generation investment modelling for government and have better access to resources than the academic sector (e.g., [104]). It is not just input data that is helpful in a model such as this, but also data about processes, in particular, how firms carry out investment decisions. Some details can be obtained by conducting interviews, but by and large this information is kept secret and so must be estimated.

- Owing to the commercial nature of energy markets, limitations exist on the amount of data that can be obtained. For instance, detailed information about the **technical characteristics** of existing power plants, such as thermal efficiencies and forced outage rates is not published, which makes calibration of the NG Winter Outlook estimates difficult.

- **Electricity demand** data for GB is readily available as far back as 2001.

All forms of recorded data can be subject to errors. Missing or corrupt measurements can lead to systematic errors in estimations. Therefore an important part of any process which involves a lot of data gathering is quality assessment. Another issue surrounds data consistency. The gathering of data inevitably leads to inconsistencies between resources (e.g., Fig. 3.5). This is particularly problematic when comparing model output with historic trends; it may match one data set well but not another.

5.1.5 System dynamics

Modern control theory operates in the time domain and differential equations involving *state variables* are used to represent the system. In this instance, the dynamics of the system state concerns the evolution of installed generation capacity and rate of demand growth. The rate of change in capacity at a particular time-step is dependent on new plant coming online or being de-mothballed together with any plant being retired or mothballed. Both are delayed signals from some earlier time. In the case of new plant this delay is the lead time for construction, and in the case of retiring plant, this delay is the design lifetime. Further delays are caused by the need for investors to accumulate data in order to form their expectations about future market conditions. Mothballing requires zero delay, whilst de-mothballing requires a delay that is significantly less than full construction. An aggregate approach is taken whereby capacity is combined into five categories of generator technology, namely nuclear, coal, wind, CCGT and OCGT each with its own financial and technological characteristics. In this implementation, all generator types are endogenous with evolution depending largely on the investment decision element of the model. Demand is modelled as exogenous using predefined growth scenarios. The model does not consider the transmission network and is conceptually a single bus system.

The elegant notation used in [157] is employed to define the state of the system at “real-time”. The state of the dynamic system holds crucial information and is defined by the vectors (I_x, ψ_x^B) where I_x is the installed capacity of plant type x and

$$\psi_x^B = \left\{(\eta_j^B, \xi_j^B), j = 1 \ldots \omega_x\right\}, \tag{5.2}$$

is the vector of new builds (superscript B). ξ_j^B is the size of investment block and ω_x is the number of capacity blocks of type x under construction. η_j^B indicates the time at which the decision is taken, with $-\tau_x < \eta_1^B < \eta_2^B < \ldots < \eta_{\omega_x}^B$, where τ_x is the expected build time and j is the total number of non zero investment years since time zero. The ordering convention ensures that the oldest investment block of size ξ_1^B will be completed first at time $\eta_1^B + \tau_x$ [157]. For instance, $\psi_x^B = \left\{(1_1^B, 10_2^B), (4_2^B, 10_2^B), (7_3^B, 10_3^B)\right\}$ implies that investments of size 10 units where undertaken at time-steps 1, 4 and 7. The control system is now defined as the volume of installed capacity $I(t)$, which is a parallel cascade of the five technology categories (nuclear, coal, wind, CCGT, OCGT). Each single category is defined by a Delay Differential Equation (DDE), i.e., the change in installed capacity at time t is given by:

$$dI_x(t) = \sum_{(\eta_j^B,\xi_j^B)\in\psi_x^B} \xi_j^B \delta(t - \eta_j^B - \tau_x) - \sum_{(\eta_j^B,\xi_j^B)\in\psi_x^B} \xi_j^B \delta(t - \eta_j^B - \tau_x - \alpha_x), \quad (5.3)$$

where $\delta(t)$ is the Dirac delta function. This function has zero width and infinite amplitude with $\int_{\infty}^{\infty} \delta(t)dt = 1$, and is zero everywhere except at the origin [175]. The first term in (5.3) accounts for the addition of capacity ξ_j^B installed at time $\eta_j^B + \tau_x$. The second term accounts for deletion of capacity ξ_j^B built at time $\eta_j^B + \tau_x$ and retired at time $\eta_j^B + \tau_x + \alpha_x$, where α_x is the expected lifetime. This is described in Fig. 5.3. Reading from left to right, the diagram demonstrates how new capacity builds are represented by nonzero impulses of magnitude ξ_j^B. In this example, impulses at times η_1^B and η_4^B contribute to the change in installed capacity at time t due to one of the two Dirac delta functions in (5.3) evaluating to zero. This demonstrates how time delays due to capacity construction (block ξ_4^B) and lifetime (block ξ_1^B) are captured. Note that neither of the Dirac delta functions for ξ_2^B and ξ_3^B evaluate to zero. Therefore they do not contribute to the change in capacity at time t, however they will contribute to the measured change in installed capacity at some other time.

(5.3) can be extended to include mothballing or premature retirement of capacity. This is defined by extending to state vector to $(I_x, \psi_x^B, \psi_x^M)$, where $\psi_x^M = \left\{(\eta_i^M, \xi_i^M), i = 1 \ldots \omega_x\right\}$ is the vector of mothballed and de-mothballed capacity (superscript M). The index i provides the total number of non zero mothballing or de-mothballing years since time zero. Thus (5.3) becomes:

$$\begin{aligned} dI_x(t) = & \sum_{(\eta_j^B,\xi_j^B)\in\psi_x^B} \xi_j^B \delta(t - \eta_j^B - \tau_x) - \sum_{(\eta_j^B,\xi_j^B)\in\psi_x^B} \xi_j^B \delta(t - \eta_j^B - \tau_x - \alpha_x) \\ & + \sum_{(\eta_i^M,\xi_i^M)\in\psi_x^M} \xi_i^M \delta(t - \eta_i^M - \tau_x^M). \end{aligned} \quad (5.4)$$

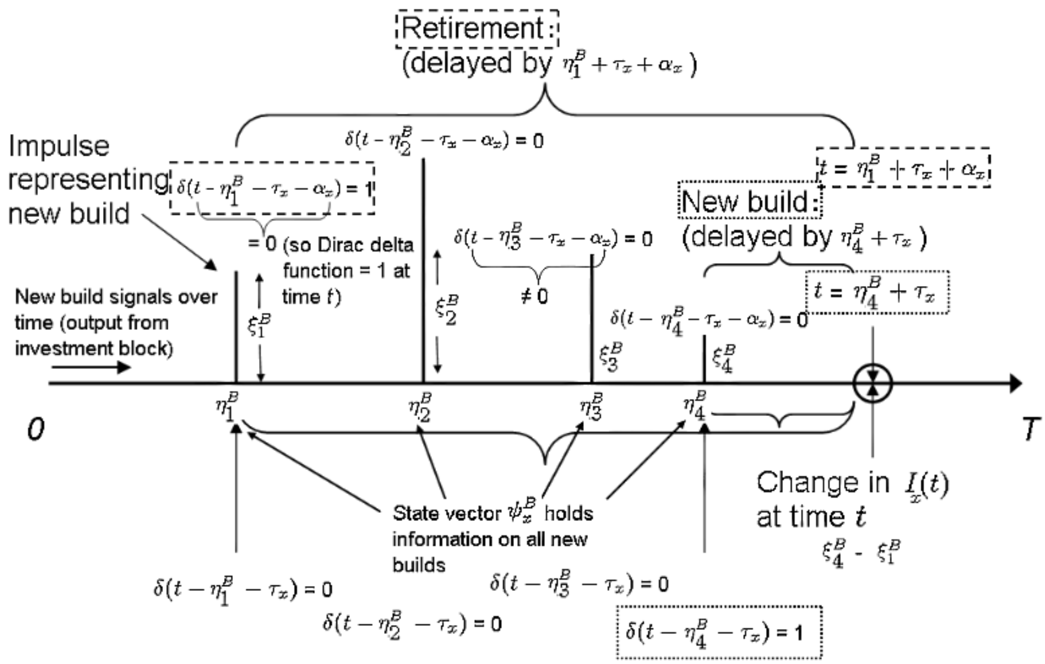


Figure 5.3: *Breakdown of DDE (5.3) at time t (indicated by a circle on the timeline) for plant type x where non zero impulses (scaled Dirac delta function) at times η_1^B and η_4^B contribute to the measured change in installed capacity $I_x(t)$.*

This includes all mothballed and de-mothballed plant because ξ_x^M is free to be positive or negative and the delay $\tau_x^M = max\left\{0, \tau_x^D \cdot sgn(\xi_i^M)\right\}$, i.e., zero if mothballing and τ_x^D otherwise, where τ_x^D is the lead time to de-mothball plant type x. Note that any mothballed capacity must continue to age whilst it is "disconnected" from the system to ensure it is permanently retired once it reaches the end of its design lifetime. It is noted that by assuming a fixed design lifetime, the model overlooks the option for generators to repower their existing plant (e.g., refurbishment, systems upgrade and equipment retrofit [176]). The implication in a model such as this is that the simulated investment dynamics may alter due to perhaps: 1) smoother exit of existing plant and 2) a lower capital intensive option for generators relative to full replacement. However representing this additional choice for investors would add significant complexity to the model and increase computational burdens. Moreover, the factors involved in a repowering decision are very complex (e.g., see [176]) and the costs are site- and technology-specific. Due to the lack of publicly available data, these costs would likely need to be estimated, and if included in the simulation model would reduce the credibility of the results presented. Each capacity block in (5.3) and (5.4) consists of a number of smaller components that represent individual plants. These are 500 MW for nuclear and coal, 100 MW for CCGT and wind, and 50 MW for OCGT.

The total installed capacity can be obtained by integrating the DDEs in (5.4)

$$I_x(t) = I_x(0) + \sum_{i \in A_x} \xi_j^B - \sum_{i \in R_x} \xi_j^B + \sum_{i=1}^{\omega_x^M} \xi_i^M \tag{5.5}$$

and summing over the five technology types, i.e. $I(t) = \sum_{x=1}^{4} I_x(t)$. $I_x(0)$ is the initial installed capacity, A_x is the set of all built plant $A_x = \left\{ j | t \geq \eta_j^B + \tau_x \right\}$, and R_x is the set of all retired plant $R_x = \left\{ j | t \geq \eta_j^B + \tau_x + \alpha_x \right\}$.

5.2 Modelling the "real-time" GB wholesale energy market

Simulating wholesale energy prices is an important aspect of any model that addresses investment in generating capacity under a liberalised market framework. Furthermore, the wholesale price trends witnessed can indicate to observers how well the system is functioning and whether there has been adequate investment in capacity. For example, as discussed in sub-section 2.4.5, there is a tendency for over-investment to lead to periods of low energy prices and conversely under-investment leads to periods of high prices. Further, investors must also make market price expectations and much of the logic presented in this section is used to anticipate future revenues for generators during the forward-looking investment decision process described in Section 5.3.

It is assumed that in a competitive energy market, generators will bid to produce electricity at or around their marginal cost. The rules of the GB market allow for generators to trade freely, i.e., there is no marginal cost bid rule in place. However there is anecdotal evidence to suggest that prices are typically close enough to theoretical marginal cost, so it appears valid to assume generators bid in this way. That said, recent analyses of the GB market [98] showed a tendency for prices in the balancing mechanism (BM) to rise above the estimated marginal cost of the last generator dispatched in peak demand hours. Other examples of empirical analyses which support this claim include [177, 178]. In [177], an analysis of the Texas BM revealed evidence of bidders, particularly smaller firms, submitting supply curves in excess of their theoretical optimal supply function. [178] found that a Cournot oligopoly model provided a better representation of the California, Pennsylvania-New Jersey-Maryland (PJM) and New England wholesale markets than a perfectly competitive model during peak hours, and it was during these hours that the presence of price mark-ups was detected. In this case the market does not necessarily clear at the marginal cost of production. Consequently, it was decided to include a price mark-up function, which is driven by the system capacity margin. This is also consistent with the Häni model. This also agrees with the methodology used in [98] to simulate GB system marginal prices. Details of each element together with how the full wholesale price is simulated is covered below.

5.2.1 The marginal cost element

This is computed using the aggregate system supply curve under the assumption of perfect competition. More precisely, variable operating costs are estimated for each of the five technology types and an aggregate supply curve based on the entire generation set is constructed.

5.2.1.1 Fuel

Firstly, wind variable operating costs are assumed to be zero. In the case of the four thermal generator types, variable operating costs are based on the common (convex) cubic function of power output [179]:

$$C(P) = a + bP + cP^2 + dP^3 + V. \tag{5.6}$$

where a is the no load cost in $MBTu$, b is the average heat rate ($MBTu/MWh$), c and d are incremental heat rates, and V is the variable operation and maintenance cost (£/MWh). To keep things simple, the cubic function (5.6) is simplified. More precisely, no start up or no-load costs are considered and linear variable operating costs are assumed for all generators, i.e., total variable operating costs for generator type x are given by $C_x(P) = b_xP + V_x$, which leads to a stepped aggregate supply function in the short-run (if bids are assumed to equal generator marginal cost, $SRMC_x = b_x$) based on the marginal costs of the entire generation set. b_x can be derived from the generator type thermal efficiency, ν_x ($\in [0..1]$) and fuel cost, F (£/MWh). Note that the fuel costs must be expressed in standardised units, so the b in (5.6) is different from the b_x used to estimate total variable operating costs, though using the fact that there are 3.413 MBTu in a MWh, either form can be derived.

Note that the following conversion factors are used when converting fuel prices into costs per unit of energy produced: 1 GJ = 9.48 therms (for gas), 1 MWh = 3.6 GJ = 3.413 MBTu [180]. For coal, the rate of conversion used from £/tonne to £/GJ is 25.12 (i.e., £100/tonne is equivalent to £3.98/GJ) [181].

5.2.1.2 Emissions

To account for the Emission Trading Scheme (ETS) in operation across Europe since 2005, the additional cost incurred by generator types who produce GHG emissions must be accounted for. This is estimated by altering the variable operating cost function:

$$C(P) = b_x P + V_x + \frac{\vartheta_x \cdot F_{car}}{\nu_x} \tag{5.7}$$

where ϑ_x is the carbon produced by burning the particular fuel type at 100% efficiency (kg/MWh) and F_{car} is the ETS (or carbon) price (£/kg). The emissions output from each fuel source is given by: gas - 185 (kg/MWh) and coal - 330 (kg/MWh) [180]. Therefore based on the thermal efficiencies in Table 5.6, CCGT generator types emit 349 kg/MWh, OCGT 474 kg/MWh and coal 943 kg/MWh. These figures are broadly in line with DECC estimates of 910 kg/MWh for coal and 400 kg/MWh for gas [182]. All other technology types are modelled with zero carbon emissions. Using these conversion factors together with the thermal efficiencies in Table 5.6, allow standardised formulas for marginal costs (£/MWh) for the five conventional plant types to be derived. These equations, along with a numerical example, are given in Appendix A.2.

5.2.2 Price mark-up

In this study the market price includes an additional *mark-up* term that alters the shape of the aggregate supply curve as the system approaches scarcity. Some studies (e.g., [105]) have found the presence of negative mark-up in very low demand hours. One reason for this phenomenon in the current GB system could, for example, be on account of CHP plants needing to continue (or start) generating because of heat commitments. However such a function is not considered here. Two forms of *mark-up* function were considered; an exponential (variation of [122]):

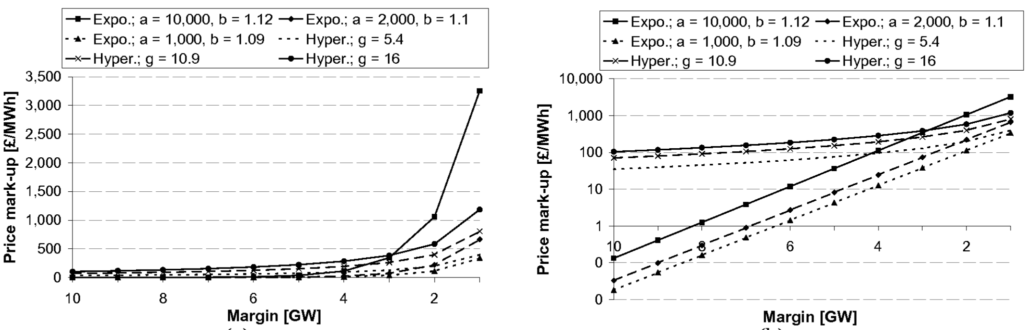
$$w_1(L, G_N^*) = ae^{b\cdot(L-G_N^*)} \tag{5.8}$$

and a hyperbolic [123]:

$$w_2(L, G_N^*) = \frac{g \cdot L}{G_N^* - L} \tag{5.9}$$

where a, b, and g are scalar factors to be set in order to induce the desired price behaviour and L and $G_N^* = \sum_{i=1}^{N} G_i$ are the load and total realised available capacity from all generators, respectively (GWs). Note $N = 5$ in this application. Note that the parameters a and b are calibrated so that a capacity margin of zero provides a mark-up equal to the VOLL. For instance, for a VOLL of £10,000/MWh, $a = 10,000$ and $b = -1.123$, and for £2,000/MWh, $a = 2,000$ and $b = -1.101$. Note that both (5.8) and (5.9) are set to zero for negative capacity margins; the situation when there is inadequate available capacity to meet load is dealt with in (5.11).

To illustrate, a selection of functions are plotted in Fig. 5.4. Both (5.8) and (5.9) were used in the Häni model [123]; results showed dependence of the model output on the function scalar factors. As it is impossible to determine the "real" pricing mechanism for the GB market, it seems sensible to tune the pricing mechanism to produce credible results for the GB market. This is the subject of sub-section 5.5.5.



(a) (b)

Figure 5.4: *(a) Price mark-up for different values of capacity margin (1 GW intervals) and calibrations for (5.8) (labelled "Expo.") for $a = 10,000$, $2,000$ and $1,000$ and (5.9) (labelled "Hyper.") for $g = 5.9$, $g = 10.9$ and $g = 16.0$. (b) Same plot on a logarithmic scale.*

The hyperbolic function was used under reasonably healthy system conditions when an adequate capacity margin is present (range 50-20%) and the risk is low, yet it does not increase sufficiently quickly as the system approaches scarcity. Under these conditions, the exponential function is better. Therefore a third *mark-up* function was constructed as a max of (5.8) and (5.9):

$$w_3(L, G_N^*) = max\left\{w_1(L, G_N^*), w_2(L, G_N^*)\right\} \tag{5.10}$$

5.2.3 Price setting procedure

The price is set by performing an economic dispatch and the marginal cost (5.7) of the last generator to be dispatched sets the system marginal production cost. By assuming a linear cost curve, the speed of the economic dispatch optimisation model is greatly increased; the supply curve is simply a piecewise step function where each piecewise constant segment is given by the SRMC of each generator type x. The full pricing function is defined as:

$$\pi(L, G_1, G_2, \ldots, G_N) = mc(L, G_1, G_2, \ldots, G_N) + w(L, G_N^*) \tag{5.11}$$

where $mc(L, G_1, G_2, \ldots, G_N)$ is the marginal production cost of meeting the load, L, given realised total available generation G_N^*, and $w(L, G_N^*)$ is the price mark-up function. If there is inadequate available generation to meet load, i.e., $L > G_N^*$, then (5.11) is set at the VOLL. The time-step in the "real-time" energy market can be lengthened to a minimum frequency of hourly. This is included so that wholesale price movements can be simulated throughout the year, which is particularly useful for model validation.

5.3 Modelling investment decisions

In this implementation, investment decisions are taken annually. Market conditions in the "real-time" energy market simulation are taken as initial conditions for predictions of the future state of the system during the lifetime of a potential investment. Crucial uncertainties such as future demand, fuel prices and wholesale energy prices are all simulated. There would be little value in listing each difference between the investment model logic in [123] and this work. It is important to note that some similarities do remain, but overall the logic has changed substantially. For instance, IRR is not used for the investment decision and the specifics of the NPV calculations have been altered. This was motivated by a need to avoid the pitfalls of IRR [36] and to allow the capacity factors of investments to be dynamic.

5.3.1 Investment logic

As Fig. 5.1 suggests, the investment block is key to determining the response of the dynamic system to changes in price and underlying levels of capacity. The arrows feeding out of the investment block to *"forward-looking models"* represents any forecasting the investor carries out in order to determine an investment strategy (i.e., invest in new capacity, withdraw existing capacity or do nothing). The "real-time" model can have up to a hourly time-step. Owing to computational burdens, the investment block executes with minimum frequency one year, with mothballing decisions carried out up to every six months. In this model the annual investment decision is considered an irreversible decision. It is based on the Net Present Value (NPV) of anticipated future profits. These profits are calculated in the standard way, that is profit is revenue received from selling power minus the costs incurred producing it.

For simplicity, a single investor who is well acquainted with the structure of the market and capable of securing the necessary debt to finance large-scale capacity investment is modelled. This representative agent approach has been used by other dynamic capacity market models (e.g., [54]). Furthermore, this design follows the adaptive expectations hypothesis discussed earlier and has been applied in other dynamic models (e.g., [34, 119, 121, 135]). When estimating the profitability of an investment, a Monte Carlo approach is taken to obtain a probability distribution of profitability whereby estimates for capacity already under construction (including delays), demand growth, and fuel prices are considered stochastic.

The representative (or aggregate) agent is aware of capacity already under construction at the start of each decision process. However there is uncertainty about a) the time remaining to build each unit; b) whether they will in fact make it to operation; and c) if rivals will jump in and invest in response to increases in expected profits. In the base case, generators assume all units will come online with 100% certainty, and remaining build time is stochastic. This is modelled as the sum of the expected build time (minus one year) plus a random variable that is sampled from a lognormal distribution, $lnN(\mu, \sigma^2)$, where μ and σ are the mean (one year) and standard deviation (six months) of the distribution, respectively. This distribution is shown in Fig 5.5. A lognormal distribution was chosen due to the fat-tailed property not present in other more "well-behaved" distributions commonly used to sample random variability about a mean (e.g., a Normal distribution). This captures investor uncertainty about additional (e.g., planning and licensing) delays: on average they last one year, though extremely long and costly delays may also be experienced. Decision making under uncertainty is modelled by taking a risk averse attitude to investment; this is discussed in the next sub-section. Investors also assume that currently operational units will generate for the duration of their design lifetime.

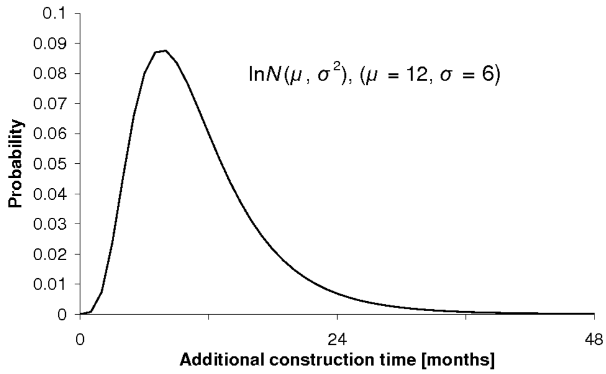


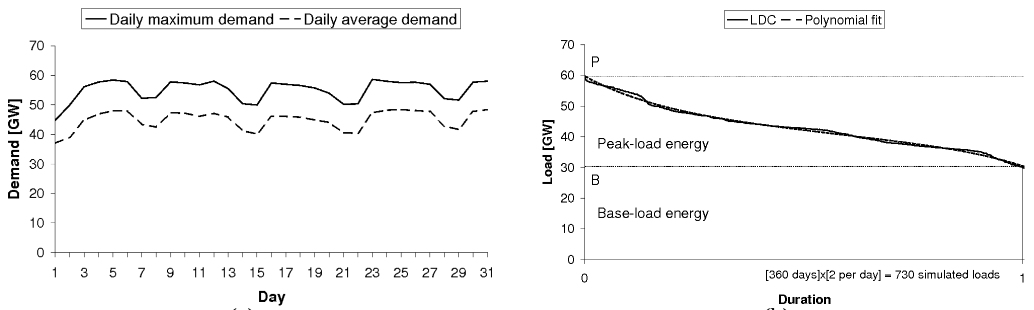
Figure 5.5: *Additional delay (months) for projects under construction. This value is added to each technology's expected build time (minus one year).*

The investor models expected price of each fuel type and the carbon price in the system as a stochastic process. More precisely a Geometric Brownian motion:

$$dF_t = m(F_t)dt + v(F_t)dW_t, \quad F_0 = f > 0, \tag{5.12}$$

is constructed for each fuel type and the parameters are defined based on perceived drift (expected change), (m) and volatility (v) characteristics. The initial values (F_0) are taken to be the average of prices witnessed over the previous year (Fig. 5.10). m is 4% per year for gas and carbon, and 1% for coal and uranium. v is 0.3% for gas, 0.2% for coal, 0.05% for uranium and 0.2% for carbon. These parameters were chosen based on anecdotal evidence about the relative volatility and rate of price increases for each commodity. Forward curves are not included in this initial iteration of the model. It should be noted that the volatility of each fuel type are not modelled as correlated random variables. The implication of this is that the model may under-estimate the relative volatility in profits between technologies, however the model is able to capture the correlation between gas and electricity prices on account of simulated electricity prices being a function of generator production costs (5.11). An example of a model that does include correlated fuel prices is [183], here a calibrated (to UK government data) SFE simulation model where oil prices are *"taken to be the underlying random variable, while the prices of gas and coal are linked to it"* is presented. The correlation coefficients used match well with empirical data for the period 1996-2005 (e.g., gas correlation with coal is 0.63). The simulation showed that predicted profits for gas-fired generators were highest and had the lowest variance, while nuclear generators profits were found to be lower and more volatile.

The yearly utilisation factor of a potential investment is based on simulated daily peak and base loads. The simulated loads use a base demand profile, which is precalculated using 2006 'IO14_DEM' load data from NG [19], plus a random variable to account for uncertainty in the forecast. This random variable follows a Normal distribution with mean 0 and standard deviation 100 MW. To model increasing uncertainty into the future, this standard deviation increases at a rate of 10% per annum throughout the forecast, i.e., $\sim N(0, 10^2 \cdot (1.1)^{y-1})$ where y is the forecast year. In addition, the investor assumes an annual growth rate of the base demand profile of 1.2%, this is applied by scaling each simulated load by 1.012. To illustrate, the base demand profile for January is plotted in Fig. 5.6.



(a) (b)

Figure 5.6: *(a) Example daily load profile (January) and (b) annual LDC (solid line) with fitted polynomial function $e(t)$ (dashed line).*

For each year in the forecast, an annual LDC is constructed from the predicted daily loads (Fig. 5.6(b)). A 5$^{\text{th}}$ order polynomial function, $e(t)$, is fitted to each LDC (using Matlab *polyfit*) and the required base-load and peak energy is computed. For simplicity, the required base-load power is assumed to be the value of the LDC at 100% duration, B, and the required base-load energy (MWh) is given by $e_B = 8760 \cdot B$, i.e., the rectangular block in Fig. 5.6(b). The required peak-load energy (MWh) is then given by $e_P = \int_0^{8760} e(t)dt - e_B$, i.e., the area above the line marked B in Fig. 5.6(b). e_B and e_P are then used to assess expected unit utilisation. Wind, coal and nuclear units are restricted to serving base-load energy and CCGT and OCGT are restricted to peak-load energy. It is noted that by forcing units to be in a particular load segment the investor does not account for 1) the fact that CCGT and OCGT units may provide residual base-load energy (i.e., remaining base-load energy once available energy from base-load units has been used), and 2) that the merit order may change over the course of the simulation time horizon. However this approach was made possible by the fact that in no simulated year in the historic GB case study was either 1) available energy from base-load capacity significantly below e_B or 2) CCGT below coal in the merit order. In an application where these factors are an issue, an alternative approach would need to be devised. Note that the improved model presented in Chapter 7 does not suffer from this imperfection.

To keep things simple, the total available energy (TAE) for each unit u (unit sizes given in sub-section 5.1.5) is calculated using expected available capacity:

$$TAE[e_u] = 8760E(G_u) = 8760c_u(1 - \rho_u), \tag{5.13}$$

where ρ_u is the unit FOR. This is subtracted from e_B or e_P following the generator type merit order (i.e., stack the units in increasing marginal cost, and call on the lowest-cost units first). The expected energy served for each unit u which is considered base-load is given by

$$E[e_u] = max(0, min(e_B - \sum_{i=1}^{N_B} E[e_i], TAE[e_u])), \tag{5.14}$$

where N_B is the total number of base-load units below u in the merit order. Similarly for a peak-load generator

$$E[e_u] = max(0, min(e_P - \sum_{i=1}^{N_P} E[e_i], TAE[e_u])), \tag{5.15}$$

where N_P is the total number of peaking units below u in the merit order. Utilisation factors can be obtained by dividing (5.14) or (5.15) (depending on generator type) by (5.13). Once the TAE from existing units has been computed and subtracted from total energy demand, any remaining energy is considered available for the potential investments (categorised as base or peak), with utilisation factors, uf_u^i, calculated similarly.

Expected competitive market prices are calculated by stacking the plants in merit order of increasing marginal cost until each predicted daily load is met. As the generators within a technology type share the same capacity and FOR characteristics, and are assumed to be subjected to independent forced outages, the expected total available generation in each period (day) is given by $E(G_N^*) = \sum_{i=1}^{N} E(G_i)$, where each $E(G_i)$ (one for each generator type) is estimated using (2.16). An example of this is depicted in Fig 5.7.

Anticipated price mark-ups are calculated by substituting $E(G_N^*)$ and the simulated loads into (5.10). The averages of the competitive peak prices (prices in those periods where load is greater than available base-load capacity) and anticipated peak mark-ups are used to calculate CCGT and OCGT gross margins (GMs, defined here as revenue minus variable cost). Similarly, the averages of the competitive base prices (prices in those periods where load is equal to, or less than, available base-load capacity) and anticipated base mark-ups are used to calculate wind, nuclear and coal GMs, although price mark-up is virtually zero in these periods due to a surplus of available resource. Thus expected GM (£) for unit u of generator type x in year i is given by:

$$GM_x^i = 8760c_x \cdot uf_u^i(E(\pi^i) - E(SRMC_x^i) + E(w_3(L, E(G_N^*)))) \qquad (5.16)$$

where c_x is the unit capacity, uf_u^i is the expected utilisation in year i, $E(\pi^i)$ is the expected system marginal price (e.g., Fig. 5.7), $E(SRMC_x^i)$ is unit u's average SRMC (calculated using annual average fuel costs from (5.12)), and $E(w_3(L, E(G_N^*)))$ is the expected price mark-up for each of the daily loads, L. Note that if $E(\pi^i) < E(SRMC_x^i)$ then uf_u^i will be zero (i.e., the unit is not anticipated to be called upon when meeting the annual load).

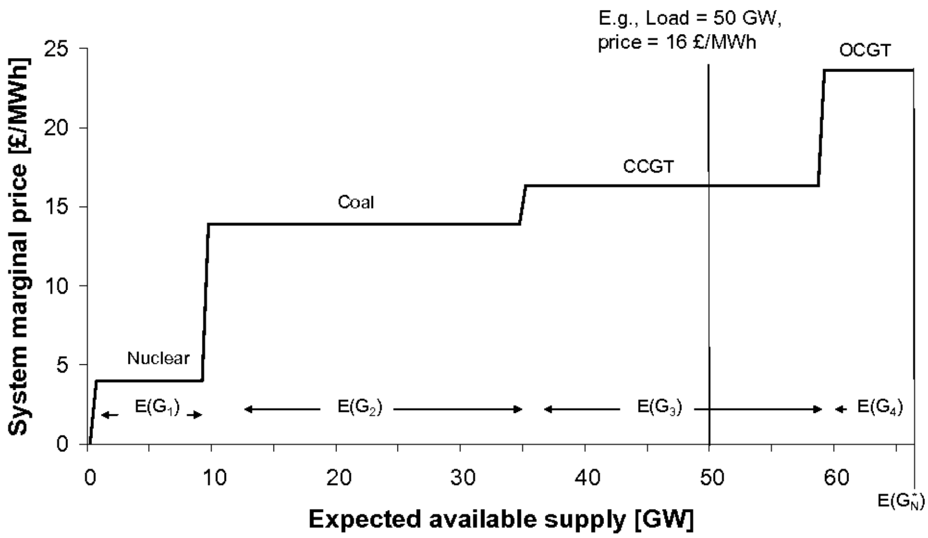


Figure 5.7: *Plot demonstrating how expected perfectly competitive market prices are calculated based on the intersection of the stepped aggregate supply function for expected available conventional generation and demand.*

This method for calculating expected output, costs and revenue is quite simplistic. It is likely to under-estimate average scarcity rent, $\pi^i - SRMC_x^i$, due to Jensen's inequality, i.e., the expectation of GMs for a given expected price, SRMC and price mark-up (5.16) is less than the GM of the expected scarcity rents, $E(\pi^i - SRMC_x^i)$, plus expected price mark-up. This methodology of calculating GMs was used for the preliminary implementation only, and a more robust approach is developed in Section 7.3.

In the case of OCGT peaking units, revenue from the STOR market is also considered. Based on the data at [184], a 2 GW annual tender with an availability price of 2 £/MW/h with a utilisation (estimated at 3% of hours) price of 100 £/MWh was included.

5.3.2 Present value of an investment

The present value of an investment in technology x at any given time is given by:

$$V_x = \sum_{i=\tau_x}^{\alpha_x} \frac{GM_x^i - FC_x}{(1+r)^{i-\tau_x}} - (IC_x + DC_x) \tag{5.17}$$

where r is the firm's weighted average cost of capital (WACC) (2.9). The expected bond return, γ, in (2.9) is set at 8% for all technologies. The remaining WACC parameters are defined in Table 5.6. GM_x^i is the gross margin (5.16) for year i, FC_x is the generator fixed operating costs, and τ_x and α_x are both expressed in years. Although the investor randomly samples capacity construction times for plant already in the build stage, for simplicity it is assumed that investors consider only fixed build times when assessing the present value of an investment (5.17). In the case of plant lifetimes, the investor assumes that capacity will close at the end of its design life. A more sophisticated representation would have the investor consider the possibility of random construction times and lifetimes, which is left for future research. IC_x is the present worth of the investment cost:

$$IC_x = \sum_{i=0}^{\tau_x - 1} M_i^x c_x p_x (1+r)^{-(i-\tau_x)}, \tag{5.18}$$

where p_x is the construction cost (£/MW), M_i^x is the capital expenditure vector for the project with $\sum_{i=0}^{\tau_x-1} M_i^x = 1$. The expenditure schedules used are given in Appendix A.3, e.g., nuclear is spread over 7 years with 10% of construction costs incurred in years 1-2, with 20% in years 3-6. Total interest accumulated during construction is given by $TIAC_x = IC_x - c_x p_x$. Finally, DC_x is the present worth of the decommissioning cost:

$$DC_x = \frac{1}{(1+r)^{\alpha_x}} \sum_{i=\alpha_x}^{\alpha_x+\delta_x} M_i^q d_x q_x (1+r)^{-i}, \tag{5.19}$$

where d_x is the cost (£/MW) of decommissioning, δ_x is the number of years required, and M_i^q is the capital expenditure vector. Only nuclear projects have considerable decommissioning costs (estimated at 12% of p_x; estimated to be between 9-12% by the World Nuclear Association [29]); in the case of other plant types the decommissioning liabilities are assumed to be offset by the salvage value of the assets [21]. Nuclear decommissioning is assumed to take 150 years and the incidence of capital outlay matrix contains 0.05 for the first 10 entries and (i.e., 50% of total decommissioning coming in first 10 years after closure) and the remaining entries are $1/140$ (i.e., 50% of cost spread over 140 years). All cashflows are discounted to the start of the first year of operation.

The total annualised (or amortised) fixed costs per unforced MW (TAFC) (£/unfor.MW/yr) discounted to the start of the first year of operation are given by:

$$TAFC_x = \frac{DCRF_x(IC_x + DC_x) + FC_x}{1 - \rho_x}, \tag{5.20}$$

where $DCRF_x$ is the deferred CRF (2.3) and ρ_x is the generator FOR.

Only the first n years of expected revenues are stochastically simulated by the investor (here $n = 15$), in contrast expected costs incurred are included for the lifetime of the plant. Investors are aware of the ROCs scheme and assume all renewable capacity (in this case wind) built after 2002 will be eligible for this subsidy.

Because the MC model randomly samples capacities, fuel prices and load growth, a large sample is required in order for investors to obtain reliable estimate of expected project value (5.17). Therefore 100 MC simulation runs are carried out for each plant type in each decision year.

5.3.3 VaR criteria

In this model, the distribution of potential profits are constructed from the MC simulations of the stochastic variables and a risk averse investor with Value at Risk decision criteria with $q = 5\%$ is assumed. The distribution of V_x in (5.17) is computed by MC simulation for each plant type and an investment is deemed attractive if $p(V_x > V_x^{opt}) \geq (100 - q)\%$, where V_x^{opt} is the minimal acceptable V_x; i.e., V_x^{opt} is a lower bound for the project value used in the VaR criterion. This is assumed to be zero as investors recover initial investment and receive adequate return on investment on account of the WACC applied in (5.17). An example of this concept is shown in Fig 5.8: the VaR for $q = 0.01$ is negative (investment is deemed nonprofitable), by comparison the VaR for $q = 0.05$ and $q = 0.5$ are both positive (investment is deemed profitable).

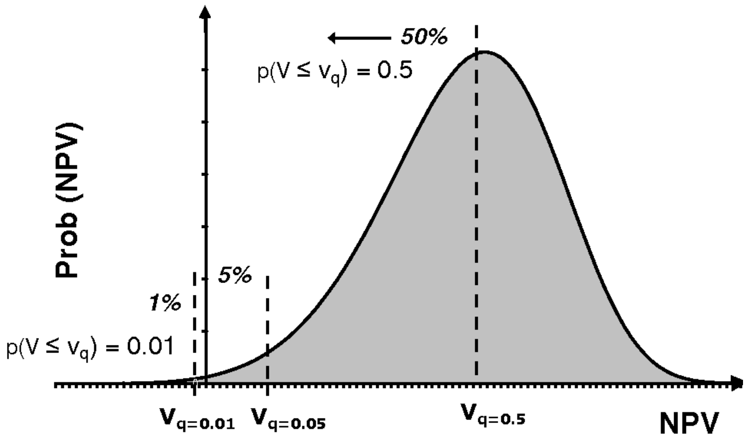


Figure 5.8: *Example of critical values, v_q, for different values of q.*

At each decision moment, the investment decision continues iteratively, whereby the decision is to 1) invest in a block of the most profitable technology (if any), followed by 2) a re-run of the investment decision with all new investments accounted for. Steps 1 and 2 are then repeated until no plant additions are profitable. There is also a secondary logical check that the expected utilisation factor for all thermal generation is greater than zero in the first year of operation ($uf_x^1 > 0$), otherwise the investment is delayed. When a number of technology types are optimal in terms of the above criteria, those projects are ranked by their Profitability Index (PI) and the option with the largest PI is chosen. The PI is defined as the ratio of project value to investment cost:

$$PI_x = \frac{V_x}{(IC_x + DC_x)} \tag{5.21}$$

i.e., critical ($q\%$) V_x divided by total investment cost.

5.3.4 Mothballing

The mothballing of existing capacity is a direct consequence of the market feedback mechanism; if a generator believes that a particular plant will not be able to cover its fixed costs from revenues in the energy-only market (and assuming it has not established a contract with the SO to provide reserve), then mothballing the plant until the price rises sufficiently is an option. It

Plant name	Unit(s)	Type	Capacity (MW)	Year
Drakelow C	9 10 12	Coal	999	2003
Grain	1,4	Oil	1350	2003[a]
High Marnham	1-5	Coal	945	2003
Killingholme	2A/B/S	CCGT	450	2002[b]
Killingholme	1A/B/S	CCGT	450	2003
TOTAL			4194	

[a]Mothballed due to low prices and de-mothballed same year (one unit in September and the other in December) in response to security of supply concerns [185]. See Fig. 5.9.

[b]Both 1x and 2x sets closed due to low prices. 600 MW of capacity at Killingholme was returned to meet the needs of a National Grid STOR [185]. Returned to full service (900 MW) in August 2005.

Table 5.1: *Plant retirements or mothballing shortly after introduction of NETA [27] (does not include nuclear retirements).*

is well documented that after the introduction of NETA, a large volume of old and inefficient plant was retired or mothballed (Table 5.1 and Fig. 5.9). This was attributed to falling wholesale prices making inefficient plant uneconomical. However some plant was retired purely based on lifetime. In line with the formulations earlier, plant closures as a result of reaching the end of design lifetime is exogenous, in contrast retirements due to economic conditions are left to the model.

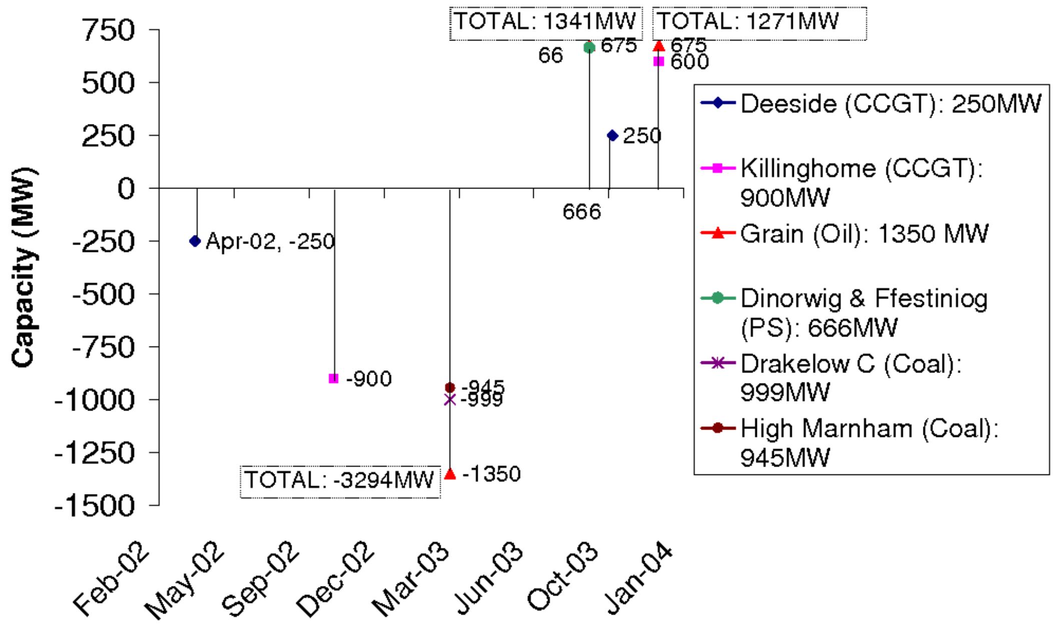


Figure 5.9: *Historic mothballing and de-mothballing in GB during 2002-4 (x-axis labels shown below plot) [173].*

Once capacity is mothballed (of which there is currently 1.25 GW in GB [66]) then it remains connected to the system but no longer contributes to short-term security of supply risk calculations (as it would take a number of months to get the plant ready to generate again from the mothballed state). It is assumed here that any mothballed capacity can be brought back at anytime up until 5 years after if was initially mothballed. If the 5 year threshold is exceeded then the capacity is permanently retired.

This model considers coal, CCGT and OCGT plant eligible for mothballing. All nuclear and wind operational fixed costs are considered sunk. The decision is taken every six months and is based on $AGM_x < 0$ where AGM_x is the average expected gross margin over fixed operating costs for the next year of operation:

$$AGM_x = \frac{1}{100} \sum_{j=1}^{100} \frac{GM_x^{1,j} - FC_x}{(1+r)}, \tag{5.22}$$

where j represents the Monte Carlo simulations (100 in total). Therefore only generator fixed and variable operating costs are relevant for this decision.

Choosing six months as the decision time-step was chosen based on historic frequency of mothballing in GB (Fig. 5.9). The iterative characteristic described above is also used, whereby plant is mothballed (or de-mothballed) until it is no longer economical to do so. Or alternatively, in an extreme case, there is no more plant left of the technology being assessed to mothball or de-mothball.

5.4 Historic GB case study assumptions

To verify the approach and model formulation, it was attempted to simulate the market dynamics in GB since the introduction of NETA in 2001. The dynamic model is applied to an ‘energy-only’ market setting with a initial capacity mix comparable to the GB power system and a VOLL of £10,000/MWh. To mirror UK policy held at the time, nuclear is not considered for investment before 2006. ROCs became active at the end of 2002 and are paid to all operational wind generation at the rate of 47 £/MWh (based on average ROC auction price in 2002 [186]).

Using available data on capacity, demand, prices and fuel costs, the model was set-up to simulate GB investment market conditions from 2001 onwards. The system capacity has been tuned to reflect the situation just prior to the introduction of NETA. The initial plant mix used the simulation is based on total transmission connected capacity in GB around that time shown in Table 5.2. To keep the model simple, minor sources of peaking capacity such as oil and pumped storage are combined with OCGT. Other base-load such as run-of-river hydro is included with coal. By taking this approach, model complexity is kept to a minimum and computational time is reduced. The capacities of each technology type is shown in Table 5.7. To get a complete GB picture, the E&W and Scotland systems are combined into a single energy market.[2]

Plant type	**Capacity (GW)**	**# stations**	**# unit sets**
Coal	30.5	19	70
CCGT	19.4	30	101
CHP	0.8	6	14
Hydro	1.1	34	8[a]
Nuclear	12.5	13	42
Oil	3.7	3	6
Onshore wind	0.12		
OCGT	1.2	19	34
Pumped storage	2.1	4	13
Total	**71.4**	**128**	**288**

[a]Independent sets (i.e., not part of same hydro scheme, cf. Table 3.3).

Table 5.2: *Approximate GB installed transmission entry capacity in 2000. Source: [27, 90].*

To reflect capacity already under construction in GB in 2001, 4.3 GW of CCGT capacity comes online during the first (1.2 GW), second (1.3 GW) and third (1.8 GW) year of the simulation.

[2]As mentioned earlier, this actually became the case in April 2005 with the introduction of the British Electricity Trading and Transmission Arrangements (BETTA) in GB.

Technology x **in (5.17)**	**Capex £/kW**	**FC £/kW/yr**	**Var. O&M £/MWh**	**Equity return** κ	**Gearing** χ	**WACC** r^n **(nom.)**	**WACC** r^r **(real)**[a]
Nuclear	1,485	43	0.18	0.15	0.5	0.115	0.089
Coal	900	21	1.01	0.12	0.6	0.096	0.069
CCGT	410	8	2.00	0.12	0.6	0.096	0.069
OCGT	300	9	2.70	0.12	0.6	0.096	0.069

[a]Assuming a 2.5% rate of inflation.

Table 5.3: *Generator financial assumptions with symbols defined in Section 5.3. Sources: [26, 39, 104].*

Station	Capacity (MW)	Final year of oper.	Model cap. (MW)
Cockenzie	1104	2011	1000
Didcot A	2108	2012	2000
Kingsnorth	1940	2012	2000
Tilbury	1104	2012	1000
Ferrybridge 1/2	980	2015	1000
Ironbridge	964	2015	1000
TOTAL	8200		8000

Table 5.4: *Assumed coal LCPD opt-out stations closure dates "under current running patterns", National Grid Winter Outlook 2009/10 [187].*

Station	Capacity (MW)	Final year of oper.	Model cap. (MW)
Bradwell	240	2003	500
Chapelcross	150	2003	500
Dungeness A	440	2006	500
Sizewell A	458	2006	500
Oldbury	470	2009	500
Wylfa	980	2011	1000
Hartlepool	1208	2014	1000
Heysham 1	1213	2014	1000
Hinkley Point B	1261	2016	1000
Hunterson	1090	2016	1000
Dungeness B	1082	2018	1000
Heysham 2	1204	2023	1500
Torness	1200	2023	1500
Sizewell B	1200	2035	1500
TOTAL	12470		13000

Table 5.5: *Assumed nuclear station closure dates. Sources [30, 188].*

Existing coal plant included in the LCPD is modelled with the reduced lifetimes shown in Table 5.4. The oil plants at Grain (1300 MW), Fawley (1000 MW) and Littlebrook D (1370 MW) are assumed to close at the end of 2015. This model assumes a fixed lifetime for all LCPD exempt and opt-in plant (Table 5.7) and that "sweating the assets" in times of tight supply conditions is not an option. Nuclear closures are consistent with those in Table 5.5. All other existing units are given retirement dates consistent with the lifetime assumptions in Table 5.6 and commissioning dates in Appendix F of [27].

Note that the build times shown in Table 5.6 are base times only; a random variable representing

Technology x **in (5.17)**	**Therm. eff.**	**FOR** ρ	**Lifetime** α **(yrs)**	**Build** τ **(yrs)**	**De-mothball** τ^D **(months)**	**Initial (GW)**	**No. of units**	**Unit size**
Nuclear	0.36	0.10[a]	40	7	-	12.5	25	500
Coal	0.35	0.14	40	5	6	31	62	500
CCGT	0.53	0.13	25	3	2	19.1	191	100
OCGT	0.39	0.10	40	2	2	8.35	167	50

[a]Recent years have shown a decline in the annual availability of the GB nuclear fleet (likely due to age), therefore this value is reduced to 75% for existing nuclear capacity. This estimate is the average of the aggregated energy output from installed GB nuclear capacity 02-09 [190]. New nuclear builds are expected 90% availability.

Table 5.6: *Generator technical assumptions and initial system capacity with symbols defined in Section 5.3. Sources: [26, 39, 104].*

Technology x **in (5.17)**	**Annualised FC £/unfor.MW/yr**	**Total interest acc. during construction £/MW**
Nuclear	244,170	474,700
Coal	111,210	173,105
CCGT	51,790	54,840
OCGT	35,555	30,655

Table 5.7: *Generator (real) total fixed and interest cost assumptions for historic simulation.*

unforeseen delays (e.g., in the planning process) is added to the build time, τ_x, after the build decision is taken (as in the Häni model [123]). Like the investment model, delays are sampled from a lognormal distribution with mean 1 year and standard deviation six months (Fig. 5.5). Furthermore, the number of projects that are abandoned is modelled using a simple discrete probability distribution such that on average $x\%$ of projects are abandoned. In the base case, $x = 0$ for coal, nuclear, CCGT and OCGT. The logic here being that traditionally, new thermal capacity is usually built on an exisiting site with a grid connection so will not come up against planning resistance. Owing to its poor record at gaining planning permission, $x = 50\%$ for wind. For instance, a basic analysis of renewableUK data [189] revealed that from 2004 to 2008, 495 applications were submitted and only 258 had been approved by summer 2009. Table 5.3 shows the costs and financial assumptions used. The total annualised fixed costs per unforced MW and total interest accumulated during construction are shown in Table 5.7.

In this implementation, wind was included in the investment decision. Its annual maximum utilisation was assumed to be 0.28, with nominal WACC, $r^n = 0.15$, capex $p = 1,000$ (£/kW), $FC = 27$ (£/kW/y), variable O&M $v = 0$ (£/MWh), lifetime $\alpha = 25$ years and build time

$\tau = 4$ years [26].

The data shown in Fig. 5.10 was used to model generator fuel and emissions prices at "real-time". The gas price is modelled with a greater degree of accuracy than the other fuel prices for a number of reasons: firstly, only detailed historical data of gas prices was obtainable, secondly by having a detailed model for gas, the characteristic of electricity prices being strongly correlated with gas prices can also be checked and verified.

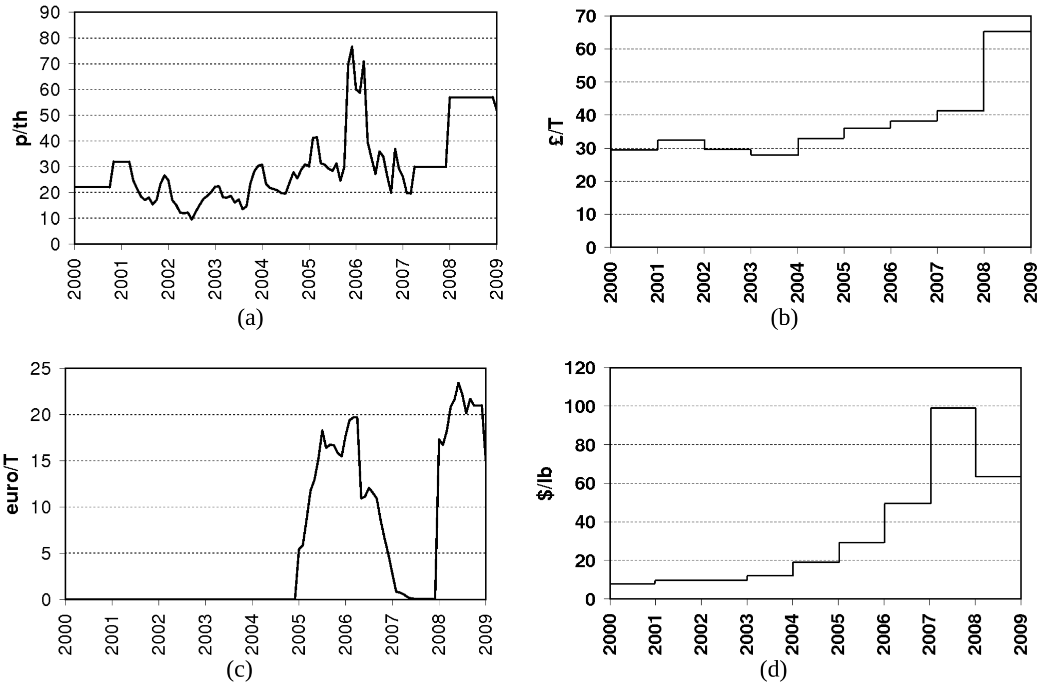


Figure 5.10: *Plots of model real-time fuel and ETS prices: (a) gas [191], (b) coal [192], (c) carbon [191] and (d) uranium [193].*

The "real-time" model can be used to test the realism of the pricing mechanism against realised prices. Therefore, the model time-step is set to hourly (investment decision frequency annual) and demand is modelled based on historic half-hourly GB demand data taken from [27]. Generally speaking, the choice of time-step in models such as this is made by considering the trade-offs between computational tractability and speed, and the granularity of model outputs; the finer the time-step the more information can be observed.

The model has been implemented using in the Matlab/Simulink 2008 environment. Using an Intel dual-core 2.40GHz processor (note that Matlab versions before 2011 will only execute using a single processor) with 3.12GB RAM with 100 Monte Carlo simulations for each investment decision and a hourly time-step simulation period of 8 years takes many hours to complete.

5.5 Simulation results

5.5.1 Base case results

Fig. 5.11 shows the model simulation of capacity margin against reality. Plainly the negative response after 2001 of the solid line is deeper than in reality but the frequency of oscillation in the two lines is very similar (which is not surprising given that the frequency of investment and peak demand is annual) and the margin also seems to be better damped than in reality. However considering the model is based on a number of assumptions and simplifications, these results show a good agreement with this data. This comparison of capacity margins is clearly motivated by the goal of using this 'signal' as the control system reference signal. Given that it is made up of both an exogenous and endogenous model parameter, i.e., demand and capacity, respectively, this match must be interpreted accordingly. For instance, the dynamics in underlying peak demand shown in Fig. 5.12 have a significant influence on the capacity margin.

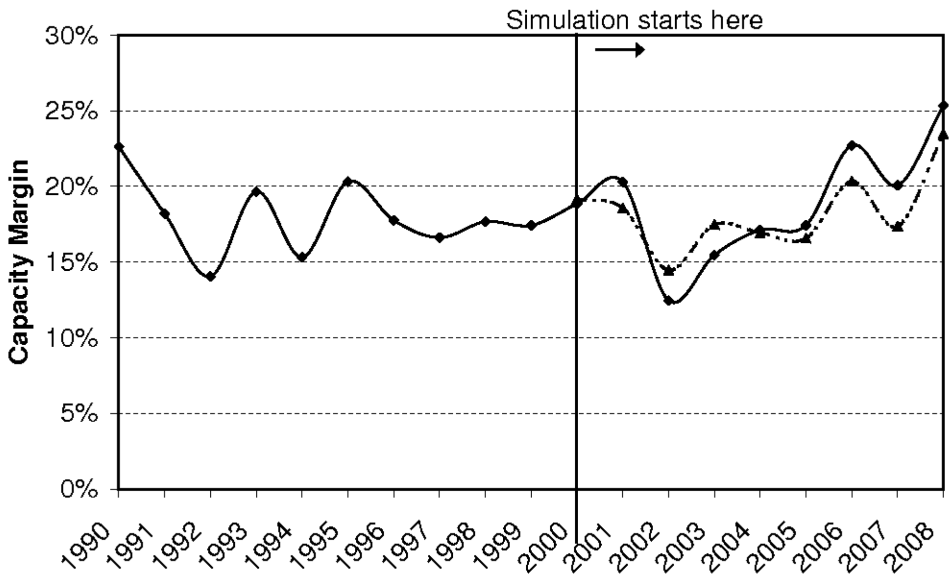


Figure 5.11: *Generation capacity margin oscillations witnessed since market liberalisation (solid line) and simulation results from 2000 (dashed line).*

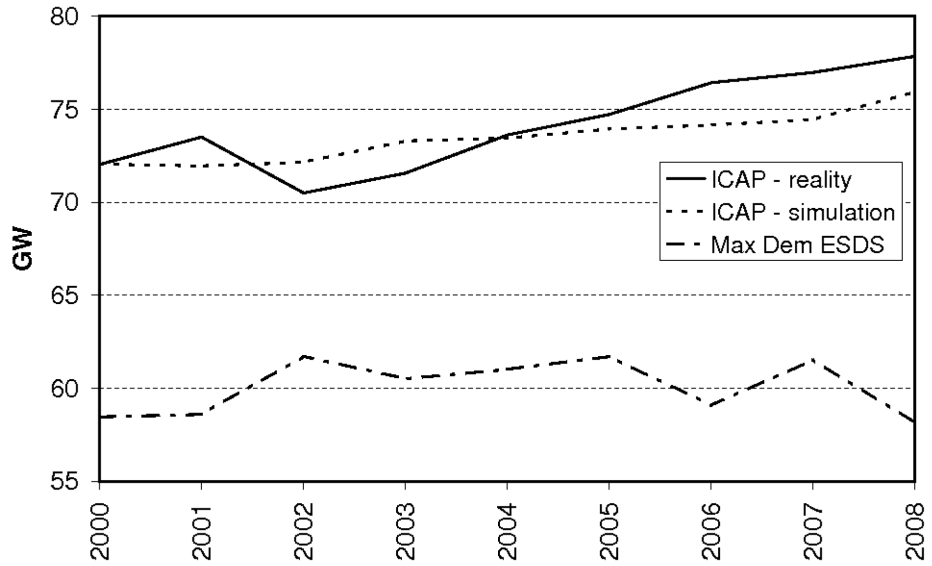


Figure 5.12: *Evolution of total installed capacity in GB in reality and in the simulation. GB maximum demand also shown.*

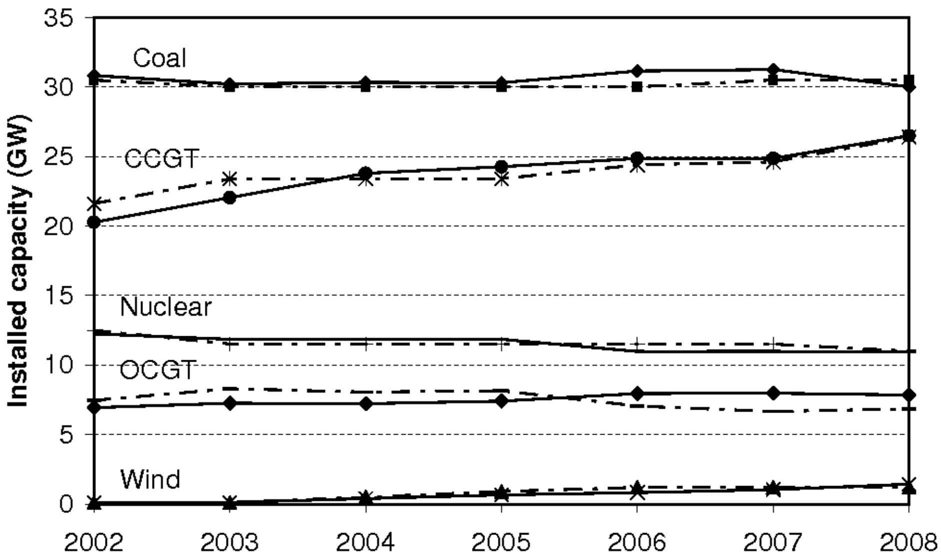


Figure 5.13: *Approximate evolution of installed capacity for each generator type in GB (solid lines) and in the simulation (dotted lines).*

Fig. 5.12 shows the evolution of total installed capacity (ICAP) in the simulation versus reality (using the ESDS [90]) in the base case (where the investment decision in annual). Both lines in Fig. 5.12 follow a similar trajectory however the simulation is smoother. Total capacity is broken down by technology types in Fig 5.13, again the match is reasonable although these comparisons depend on the approximations used (in this case according to DUKES [190]).

5.5.2 Investments

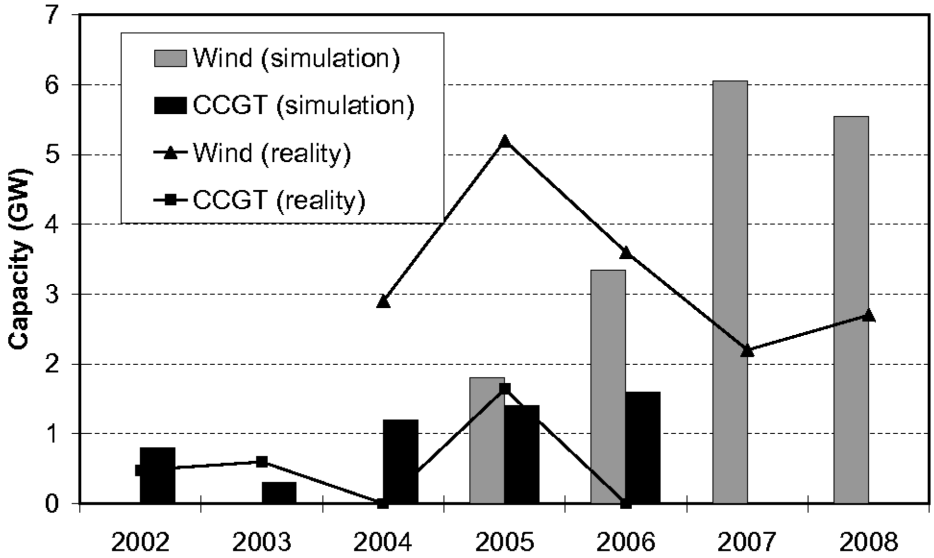


Figure 5.14: *Wind and CCGT capacity investment triggered during the simulation that actually completes the build stage, and actual investments announced in GB. Triggered wind investments have been scaled by the probability of project abandonment, i.e., 0.5. Data on wind from renewableUK [189] and CCGT builds estimated from NGC seven year statements [27] assuming 3 year build time.*

Fig. 5.14 shows the volume and type of investment triggered during the simulation. There is little investment in the early years owing to the capacity already under construction when the simulation starts (Section 5.4).

Wind power investment is also another key area of interest and was modelled as an endogenous model parameter in this case. Fig. 5.14 shows that in the simulation no significant investment in wind occurs until after 2005. After this, not only does the ROC subsidy become active, but also there are significant extra costs incurred for other projects: for CCGT the cost of gas increases by 60% from 24 p/therm to 39 p/therm and coal generators have the additional cost of emissions to consider. Fig. 5.14 also shows renewableUK data [189] on applications received for new wind farms received in GB. The main points to note here are although wind power investment is delayed until 2005, perhaps owing to the high WACC for this technology, there seems to be an over-investment in later years. Recalling that as the model does not consider grid connection arguably this trend is quite plausible in this model where wind is anticipated to be heavily subsidised and grid connection straightforward. Of course much of this investment will not make it to the build stage owing to the high probability (0.5) of project abandonment for reasons outside pure market rules.

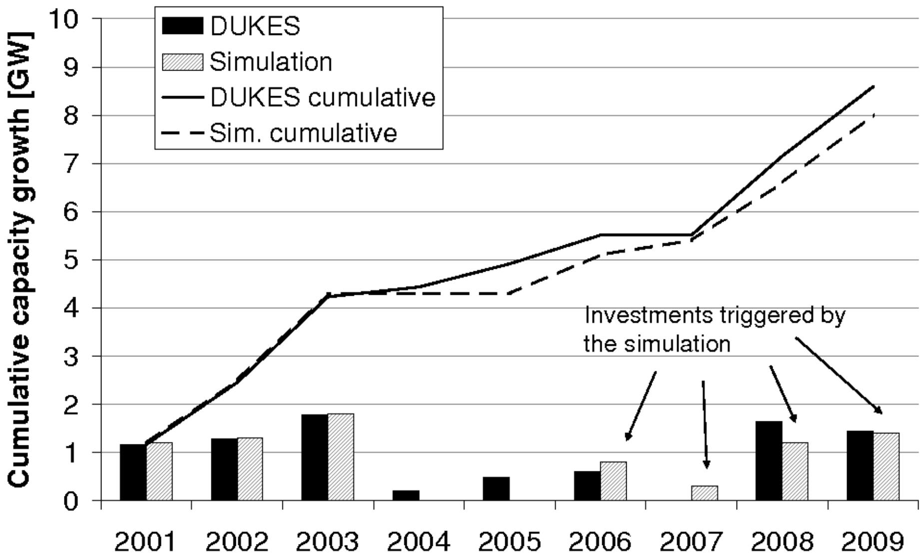


Figure 5.15: *Cumulative CCGT capacity added to the system in the simulation (dotted-line) and according to the DUKES/ESDS data [190] (solid line).*

Also of interest is the cumulative CCGT capacity added to the system, which can be see in Fig. 5.15. The new builds initialised in Fig. 5.14 can be seen coming online after the 3 year lead time (contributing to capacity in the following year). The pattern does not match perfectly, with the minor increase in 2005 not captured. By comparison the ramping up of investment in the later years, together with the relative lump sizes in, say 2006 and 2009, is encouraging, given that project lead times and precise retirement data are being estimated.

5.5.3 Mothballing

Although in reality CCGT capacity growth remains constant throughout the simulation (Fig. 5.15), the same cannot be said of OCGT plant: inspection of Fig. 5.13 shows a downturn in OCGT capacity in 2005 that does not match with reality. This downturn can be explained by a sharp increase in mothballed plant in 2005 (Fig. 5.16) in the simulation and thus the installed capacity also falls behind reality. The likely cause of this behaviour stems from a combination of an increase in installed CCGT capacity, which reduce OCGT revenues from the energy market, together with a 60% increase in gas prices. Limited information is available pertaining to when exactly GB's (current) 1.25 GW [66] of plant was mothballed. The information that was available concerned the mothballing of plant in early 2003, with much of it returning to the system in the winter of the same year (Fig. 5.9). Owing to the reduced complexity of the model compared with reality, it is difficult to capture this response because individual plant characteristics are not included in this aggregated model. Therefore although some of the preliminary simulation results were quite encouraging, the model was not able to replicate all forms of market response.

Figure 5.16: *Volume of capacity in mothballed state during the simulation.*

5.5.4 Importance of the reserve market

A sensitivity test was carried out whereby a weighting coefficient, w, in the range 0 to 1 was applied to the additional revenue received by OCGT plant for participation in the STOR market. For example, $w = 1$ represents the base case discussed above where OCGT participates in STOR, and $w = 0$ represents a complete removal of the STOR market. Fig. 5.17 shows the importance of having a separate market for reserve (which is essentially a capacity payment for OCGT plant); without it the initial capacity margin slump is deeper owing to intensified withdrawal of OCGT plant from the system. Owing to greater CCGT investment in the early years than in the base case, the margin does recover after 2005, although CCGTs are not well suited to providing reserve and hence the level of security of supply risk is greatly increased.

Interestingly the introduction of w leads to a better initial downturn in capacity at the start of the simulation ($w = 0.5$) and although the margin remains below reality in subsequent years, the difference is constant implying that this payment is a key parameter in the model. This can be interpreted as the inclusion of a steady revenue stream for peaking plant removes much of the uncertainty surrounding utilisation and price, reducing the volume of mothballed OCGT in the model.

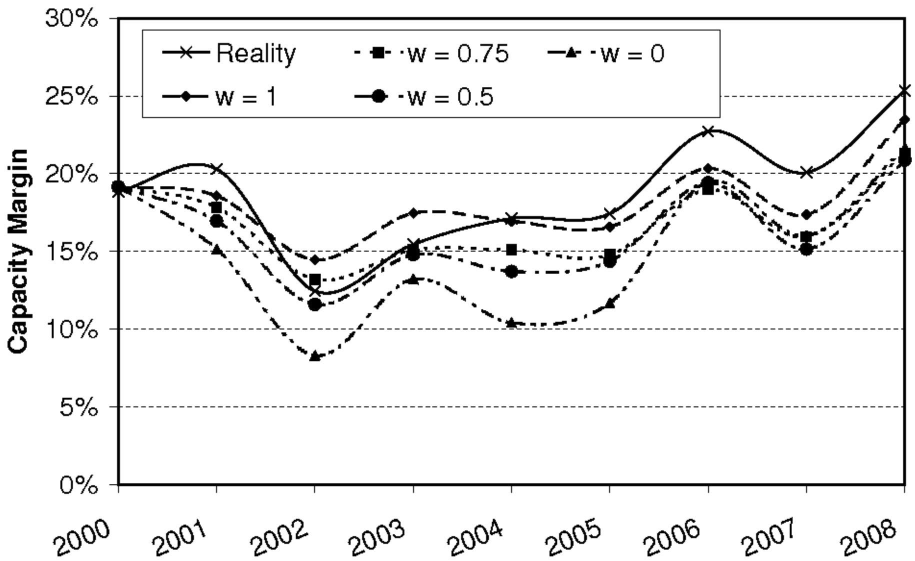


Figure 5.17: *Plot of GB generation capacity margin (solid line) with reserve market sensitivities (dashed lines) for various coefficient weightings (w).*

5.5.5 Wholesale energy prices

Another key area of analysis was the ability of the model to simulate wholesale energy prices. Using the pricing mechanism described in sub-section 5.2.3 with price mark-up (5.10), with parameters set at $a = 10,000$, $b = -1.123$, $g = 5.4$, $s = 1$ [123] (Fig. 5.4), simulated average monthly wholesale prices were compared with the monthly average forward Market Index Price (MIP, reflection of the short term wholesale price) [191]. Fig. 5.18 shows a good overall matching of trend. The loss of precision after April 2008 can be explained by recalling that detailed gas price data for these years was not available (Fig. 5.10). The ability to push through high fuel costs is demonstrated when the simulated prices for April 2001 to April 2007 are viewed in isolation. The mean difference between the MIP and simulated prices was 4.0 £/MWh with standard deviation 9.6 £/MWh. By comparison over the range using more accurate fuel price data (up to time indicated by horizontal line in Fig. 5.18) was 0.70 £/MWh with standard deviation 4.35 £/MWh.

It should be emphasised that the main aim of the model is not to predict the electricity prices but to investigate whether or not an 'energy-only' market is susceptible to excessive capacity oscillations. Hence the main aim is to investigate the dynamics of aggregate investment whatever happens to the primary fuel prices. Obviously as the absolute level of electricity prices depends strongly on the primary fuel prices, predicting future electricity prices would require a sophisticated gas and coal price prediction model and that would be beyond the scope of this research. Any changes in primary fuel prices have the effect of shifting the level of prices up or down, assuming constant GMs for gas-fired ("spark spread") and coal-fired ("dark spread") generators when selling electricity. Further, it could cause a switch in primary investment from one technology to another (e.g., from coal to gas as has been the trend in recent years). This is justified by inspection of Fig. 5.19 and 5.20 which show results from a simulation using predicted gas and coal prices from OFGEM [96] and not actual data from 2001-2007. Plainly the price graph is different (influence of demand profile is more noticeable) in contrast the capacity graph is only mildly affected.

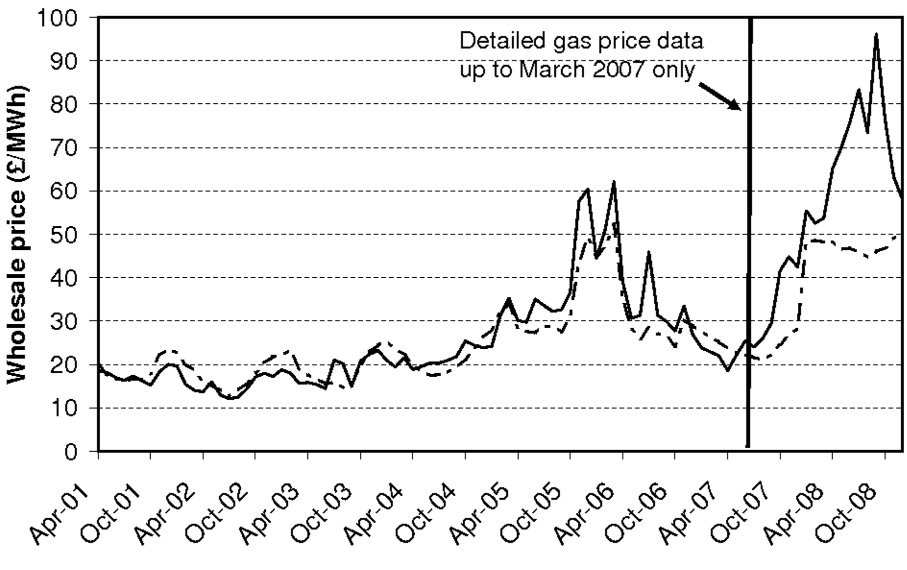


Figure 5.18: *Average monthly wholesale energy prices in the simulation (dashed line) and in GB wholesale market (solid line) [191].*

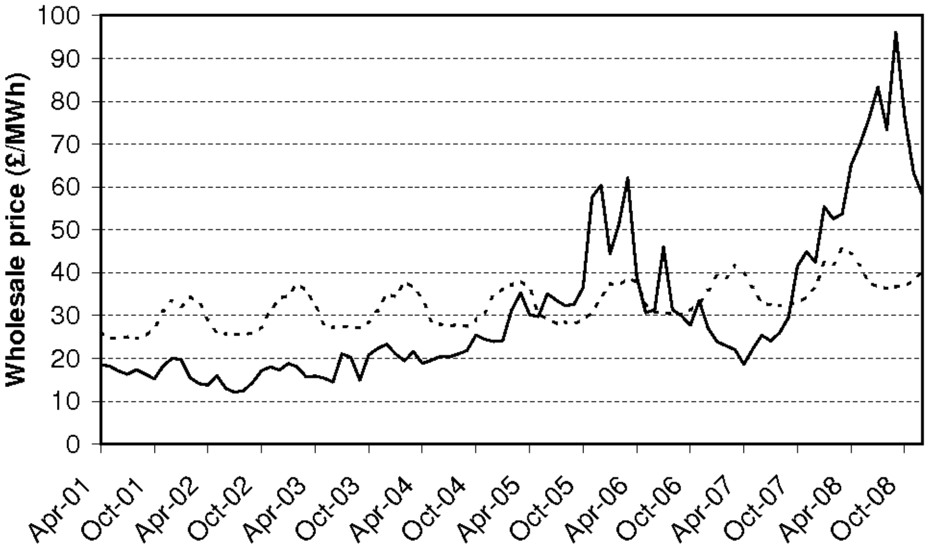


Figure 5.19: *Average monthly wholesale energy prices in the simulation when using fuel price predictions (dashed line). Actual GB wholesale market data is included for a comparison with Fig. 5.18 (solid line).*

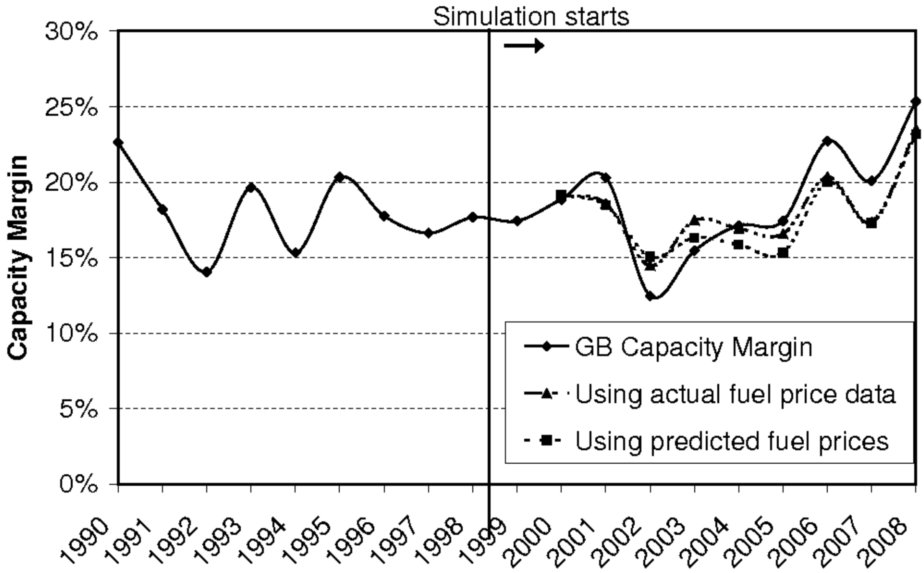


Figure 5.20: *Generation capacity margin oscillations witnessed since market liberalisation (solid line) and simulation results from 2000 with actual prices (dashed line, as in Fig. 5.11) and predicted prices (dotted line).*

5.5.6 Summary

There is no economic model that has been able to simulate past events with 100% accuracy. Attempting to replicate historic market dynamics in GB, particularly plant response to market conditions, has proved to be a challenge. The preliminary results are encouraging, although it is clear that this iteration of the model is unable to replicate the events of the GB market entirely. This is hardly surprising given the simplified nature of this model, which was necessary in order to keep the model tractable within the duration of a 3 year project.

There is scope to improve the model. It was discussed earlier that the dynamics in underlying peak demand have a significant influence on the capacity margin. Further, the results shown in Fig. 5.11 use realised peak demand. In hindsight, for the results presented here, the historic simulation should have been compared in a way that minimises the significance of this variable whilst still reflecting the security of supply risk. The realised ACS peak considers underlying demand patterns and *"typical"* winter peak weather conditions and is therefore a better measurement to use. Therefore, using the data in Fig. 3.6, the capacity margin comparison is re-plotted in Fig 5.21. While much of the oscillation is removed as a result of the use of ACS peak, the signals compare less well, with the winters 01/02 - 03/04 and 06/07 particularly poor. Also shown is the realised capacity margin using the DUKES [190] for ICAP instead of the ESDS [90], the simulation compares better with the DUKES data, with an increasing trend after 03/04 present in both signals. Furthermore, to provide more robust conclusions, a longer period of history should have been analysed, though data before the introduction of NETA could not be obtained. Data for E&W pool prices and what was believed to be aggregate E&W demand, was eventually located 3 years into the work. It could therefore not be realistically included.

That said, this stage of the work formed a vital part of the process and highlights why these (and subsequent) results should be interpreted with a degree of care. That is, replicating absolute levels of, e.g., capacity margin is difficult, yet by considering overall trends in response to different policies, participant behaviour, capacity mixes and underlying data assumptions can be considered. Furthermore, given the short duration of the validation period, where the impact of many triggered investments are not realised within the simulation time horizon, a longer simulation period must be considered to allow robust conclusions about model performance to be drawn.

As a result of feedback from the publication of the paper version of this chapter [194] and subsequent collaboration with Prof. Benjamin Hobbs of Johns Hopkins University, it was decided

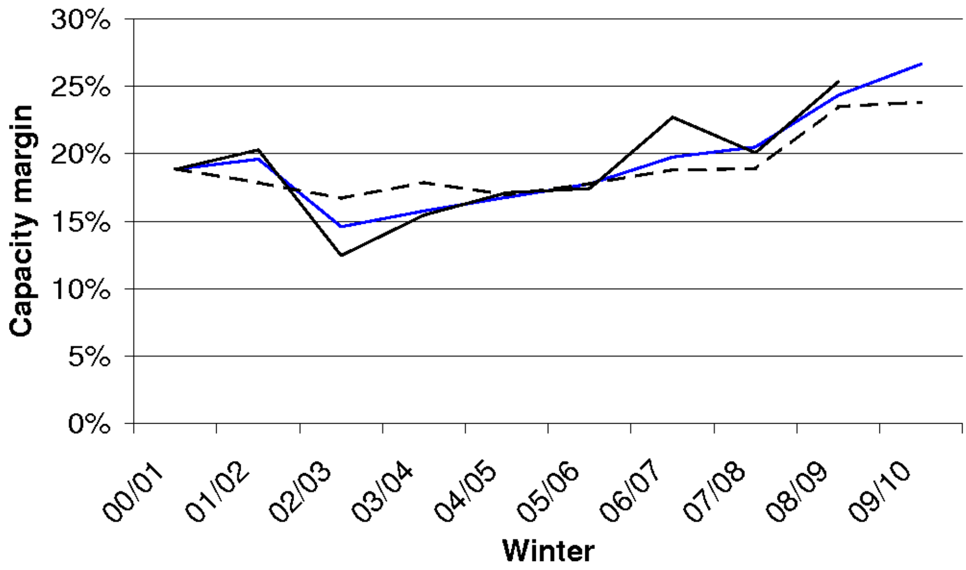


Figure 5.21: *Generation capacity margin in GB using ACS peak demand. Using historic data on ICAP from the ESDS [90] (black, solid line), and the DUKES [190] (blue line) and simulation results (dashed line) shown.*

to consider investment in wind capacity as a exogenous model parameter and focus on developing a more robust modelling technique to estimate thermal generation investment dynamics. Justification for this approach to modelling wind is the topic of Chapter 6 and the updated model, including the new method of calculating expected output, costs and revenue of thermal generation is presented in Chapter 7.

5.6 Chapter summary

In this chapter the dynamic simulation model has been presented. The basic concept of the modelling approach with a description of the feedback mechanism being investigated was given. A review of the work of Häni [123] showed that although this concept had been implemented in a simple model it had not yet been applied to an existing system. By describing the dynamics of the model mathematically and its application to the GB investment market the intended direction of the work was introduced. Next, the methods used to model generator SRMC and price mark-ups due to market power within the modelling environment were described. Section 5.3 provided details about investor logic, that is a single investor who uses NPV of expected future profits and VaR as the decision criterion. Section 5.4 described the input assumptions for GB case study, in particular the application of the model to an historic simulation of market dynamics in the period post NETA in order to achieve the goal of model verification. Section 5.5 provided the preliminary results for the model application, which showed some agreement with reality. A summary of the work demonstrated the limitations of the model and challenges encountered. This sets the scene for the next two chapters where substantial improvements to the model are undertaken in a second phase of the work.

Chapter 6

Modelling High Penetrations of Wind

Investigating how best to model increasing penetrations of wind in a dynamic investment simulation model has formed a substantial part of this work. In its own right it has provided a contribution to knowledge beyond that of the dynamic investment model work. In this chapter, the cutting-edge techniques used to model production from high penetrations of wind power are described. The results from this chapter are employed in Chapter 7 to update the investment market simulation model for assumed levels of installed wind. The work presented helps answer the following questions: 1) given historic wind resource availability in GB, what would have been the hourly production for a higher level of installed wind capacity, and 2) what is the likely contribution from wind to meeting peak demand?

6.1 Motivation

In GB, the installed capacity of wind generation is expected to increase dramatically in the coming years. Inspection of the latest renewableUK data [189] on wind projects in operation, construction, or consented there could be a capacity in excess of 15 GW onshore and up to 40 GW offshore by mid 2020, which is a significant change from today's level of $\sim$5 GW (with $\sim$2.2 GW and $\sim$2.8 GW at the transmission and distribution level, respectively [27]). This magnitude of increase in total installed wind capacity in a system with around 75 GW of total transmission connected capacity and a maximum and minimum annual demand of 60 GW and 20 GW, respectively, transforms the GB system from a low to high wind penetration. High penetrations of wind power introduce a number of challenges in power system operation and planning. There is a statistical relationship between wind availability and demand (via the dependence of both on the weather) and, unlike conventional thermal generation units whose individual availabilities are assumed independent, wind power output is related to that at other geographical locations.

6.2 Approach taken

Rather than treat wind as an endogenous parameter in the model, investment in wind capacity is now considered exogenous to the model (Fig. 5.1). This approach is justified by the fact that, to date, large-scale investment in wind capacity is driven by political, rather than economic considerations. It is therefore assumed that policies promoting investment in wind generation (e.g., ROCs) are successful in meeting renewables targets and the purpose of this work is to provide insights into the response of investment in thermal generation and subsequent levels of security of supply risk. This also seems reasonable given the results in sub-section 5.5.2 where a high level of profitability of wind led to large volumes of wind investment, thus suggesting that factors beyond pure market rules influence the growth in this technology. That said, the contribution of wind generation to meeting demand must still be considered.

In most circumstances, it is reasonable to consider wind as a “must-run” or non-dispatchable form of generation, where if the turbines are available (in the technical sense) and generating, then they will be dispatched. Network constraints may limit the amount of generation that can be exported from a particular region, and there is a large volume of work concerned with the best approach to network operation and planning in GB with large penetrations of variable wind (e.g., [68, 195]). In contrast because this work conceptually uses a single bus system, these issues are beyond the scope of this study. As a result, the residual load for thermal generation is of interest. The residual load is simply the expected load minus the expected output from wind power at each hour. By taking this approach, the analysis naturally takes into account spatial relationships in wind availability and its relationship with demand.

Measured wind output data is hard to obtain in GB and even when available (e.g., aggregate transmission connected wind output is available from bmreports [102]), these measurements do not fully represent the long-term contribution of wind, particularly offshore. Therefore, a simulation of wind output, or more precisely aggregated wind capacity factors (CFs), is the approach chosen. Two similar methods are employed when simulating onshore and offshore wind CFs, the details of which are described in the following sections.

6.2.1 Onshore wind (after [61])

Simulated hourly onshore wind output is calculated using the technique described by Olmos [61]. The Olmos study takes hourly UK Meteorological Office wind speed data[1] for 2001-07 and firstly extrapolates to adjust for hub height (10m to 60m), and secondly performs a transformation to account for the turbine design (Bonus 2 MW wind turbine power curve). Aggregate hourly GB CFs are simulated by taking a weighted average of regional CFs, based on the aggregated wind capacity in operation, construction, or consented in each region [197]. This calculation is summarised by:

$$GBCF = \frac{\sum_{i=1}^{R} w_i x_i}{\sum_{i=1}^{R} w_i} \tag{6.1}$$

where w_i are the weights, x_i are the regional CFs and R is the total number of regions. These regions are shown in Fig. 6.1 and weightings in Table 6.1. This approach is justified by the fact that although the current penetration of wind generation in GB is small relative to long-term expectations, the geographical dispersion is broadly in line with “long-term” expectations for each region.

In this study the same approach is applied to wind speeds extracted for 2005-09. The reason for choosing these five years to perform the analysis is to compliment the data that was available from a unique offshore simulation model described next. The main justifications for using wind speed measurement data when simulating aggregate onshore wind production are: 1) there is the lack of extensive time series data with adequate temporal resolution for currently operational farms in GB; and 2) the amount and geographical dispersion of onshore wind farms will grow over time, and will include production from regions where CFs and correlations with loads and wind from other areas are not currently known.

As with any expansive data set, there were issues with missing or corrupt data points. Although minimal, those that were missing or erroneous were filled by simply taking the average of the two adjacent readings in the time series. A Vestas V80-2 MW power curve is used [198] with a transform from 10m to 80m using the classic power law equation [61]:

$$\frac{v_z}{v_{z_r}} = \left(\frac{z}{z_r}\right)^{\alpha} \tag{6.2}$$

[1] In fact, the wind speeds are the mean wind speed over 10 minutes measured from minute 40 to minute 50 of the hour [196].

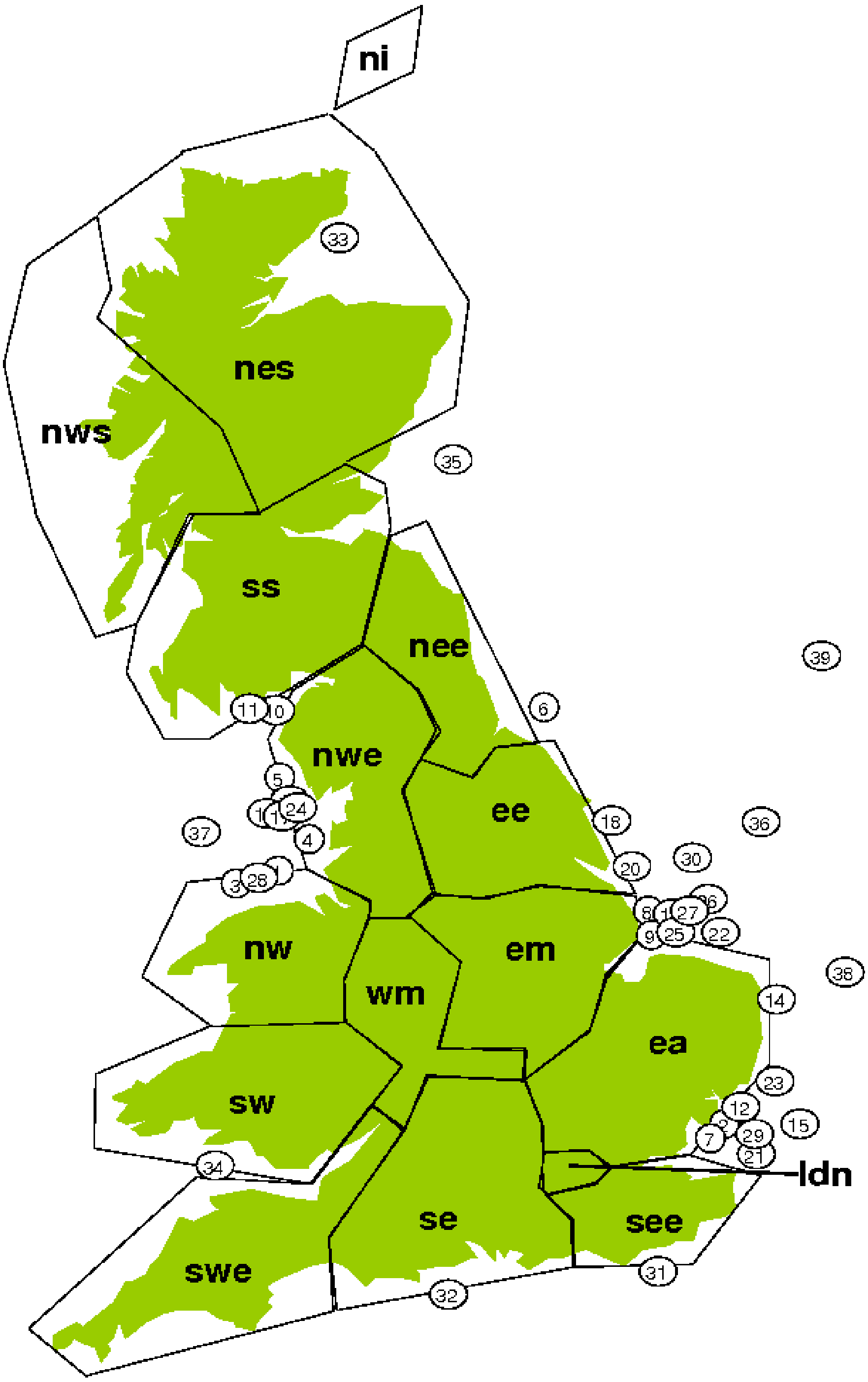


Figure 6.1: *Diagram showing onshore regions used when simulating aggregate GB CFs. Ovals positioned offshore indicate location of farms listed in Table 6.2.*

where v_z is the wind speed at height z, v_{zr} is the reference wind speed at height z_r and α is the power law (or wind shear) exponent, commonly assumed to be $\alpha = 1/7$ [199]. Note that issues concerning diurnal variation of expected onshore wind output derived from wind speed, versus wind power derived from real wind farm data are not considered here. For instance, it is necessary to decrease the wind shear exponent in (6.2) when it is sunny, because the boundary layer becomes more turbulent [200].

Region	**Weight**
ea	331.0
ee	482.3
em	423.3
nes	2319.3
ni	3.7
nw	1599.4
nwe	1389.7
nws	915.3
se	95.9
see	90.9
ss	4394.7
sw	517.8
swe	265.0
wm	30.4
TOTAL	13537.3

Table 6.1: *Onshore regional weights (MWs).*

6.2.2 Offshore wind

For GB offshore wind, the expected hourly CFs are calculated using the wind speed data from the fully compressible, non-hydrostatic mesoscale weather forecast model developed at The University of Edinburgh [201, 202]. This is a atmospheric model similar to the Met Office Unified Model [203]. The 3 km resolution (other studies have used a much lower resolution, e.g., [98] used wind speeds at a 12 km resolution) provides a good level of detail in the resource approximation. The simulated wind speeds have been validated against measured offshore wind speed data for a number of sites [204], and in the absence of extensive measurement data, provides credible estimates in time and space for the GB offshore wind resource. Simulated hourly data (average of 10 minute resolution wind speeds during the hour) was extracted for 2005-09 and the offshore locations displayed in Table 6.2. These are displayed in Fig 6.1 as numbered ovals.

The simulated wind speeds are transformed to site level CFs using the same techniques as [61] but instead using a Vestas V90-3 MW power curve and a hub height of 100m (on account of the relative increase in turbine size witnessed offshore). The weighted offshore GB CFs (6.1) are then calculated based on the results from the Round 1 to 3 Crown Estate Round auctions [197]. This data is displayed in Table 6.2; the weights w_i are displayed in right-most column, x_i are the site (not regional averages as for onshore) CFs and $R = 39$.

This approach takes a "long-term" view about the offshore wind resource. More precisely, a fully diverse offshore wind resource is available for each level of installed capacity (i.e., each additional MW is spread evenly across all sites). Using the long-term aggregate CFs may overestimate the contribution from offshore wind in the years leading up to 2020 (when large round 3 sites expected to come online). Accounting for all the issues in a model such as this would take a lot of time to implement and may not add sufficient value to the results in light of simplifications made elsewhere. It has been examined in a preliminary manner in a recent paper [204].

6.3 Calibration

Simply scaling 10m wind speed to hub height and transforming to CF using a manufacturers' power curve without calibration will probably not give a realistic long-term CFs on account of inaccuracies in scaling, turbine outages and topographic effects. For instance, wake losses and gusts (particularly for wind speeds approaching full power output) are not accounted for and so some form of calibration against measured power outputs must occur.

The approach taken here is to scale the aggregated annual CFs for onshore and offshore wind to match published data on annual capacity factors [190]. More precisely, a minimisation of the sum of squares of the residuals, i.e, the difference between observed CFs and simulated CFs (averaged across the year), is used to obtain the scaling factors. This was computed using the MS Excel 2003 Solver add-in. This method is reasonable for onshore wind because the diversity of the current resource is fairly reflective of the future (Table 6.1 and Fig. 6.1). In contrast this is not the case for the current offshore wind resource. The 1.3 GW of offshore wind currently operational [197] in GB does not fully represent the future geographical diversity of

Map No.	Name	Auction round	Turbine type	Turbine capacity (MW)	No. of turbines	Total MW w_i's
1	**North Hoyle***	1	Vestas V80	2	30	60
2	**Gunfleet Sands II**	1	Siemens	3.6	18	64.8
3	**Rhyl Flats**	1	Siemens	3.6	25	90
4	**Burbo Bank**	1	Siemens	3.6	25	90
5	**Barrow***	1	Vestas V90	3	30	90
6	Teesside	1		3	30	90
7	**Kentish Flats***	1	Vestas V90	3	30	90
8	**Lynn - Inner Dowsing**	1	Siemens	3.6	27	97.2
9	**Lynn**	1	Siemens	3.6	27	97.2
10	**Robin Rigg - East**	1	Vestas V90	3	30	90
11	**Robin Rigg - West**	1	Vestas V90	3	30	90
12	**Gunfleet Sands I**	1	Siemens	3.6	30	108
13	Ormonde	1		5	30	150
14	**Scroby Sands***	2	Vestas V80	2	30	60
15	Greater Gabbard - Galloper	2		3.6	40	144
16	Walney I	2		3.6	51	183.6
17	Walney II	2		3.6	51	183.6
18	Westernmost Rough	2		3	80	240
19	Lincs	2		3.6	75	270
20	Humber Gateway	2		3.6	83	300
21	**Thanet**	2		3	100	300
22	Sheringham Shoal	2		3.6	88	315
23	Greater Gabbard - Inner	2		3.6	100	360
24	West Duddon	2		3.6	160	500
25	Docking Shoal	2			100	540
26	Dudgeon	2			168	560
27	Race Bank	2		5	88	620
28	Gwynt Y Mor	2		3.6	160	576
29	London Array I	2		3.6	175	630
29	London Array II	2			166	370
30	Triton Knoll	2				1200
31	Hastings	3				600
32	West Isle of Wight	3				900
33	Moray Firth	3				1300
34	Bristol Channel	3				1500
35	Firth of Forth	3				3500
36	Hornsea	3				4000
37	Irish Sea	3				4200
38	Norfolk	3				7200
39	Dogger Bank	3				9000
TOTAL						40759.4

Table 6.2: *Assumed offshore wind farm developments based on data from Crown Estate auctions [197]. Sites highlighted with * indicate technical availability data has been located [205].*

the system (Table 6.3). Therefore scaling the simulated CFs by the published annual CFs would ignore increases in resource availability as a result of diversification. Consequently, the scaling factor is applied to the simulated aggregate CFs that include operational (at the time of writing) offshore farms only. These locations are highlighted in bold font in Table 6.2. Thus the GB CFs are computed for the existing 13 sites ($R = 13$ in (6.1)) and the annual average CFs from this time series are used to calculate the scaling factor. This gives a scaling of 0.928 for onshore and 0.918 for offshore. The results of the scaling are presented in Table 6.3.

	2005	**2006**	**2007**	**2008**	**2009**	**Long-term average**
DUKES (Table 7.4)						
Onshore	0.264	0.272	0.275	0.270	0.274	
Offshore	0.272	0.287	0.256	0.304	0.260	
Simulated GB CFs (unscaled)						
Onshore	0.297	0.282	0.288	0.301	0.290	
Offshore (existing farms)	0.301	0.299	0.288	0.320	0.268	
Offshore (long-term)	0.474	0.447	0.457	0.495	0.442	
Scaling factor onshore	0.928					
Scaling factor offshore	0.919					
Simulated GB CFs (scaled)						
Onshore	0.276	0.261	0.267	0.280	0.269	0.271
Offshore (existing farms)	0.277	0.275	0.265	0.294	0.247	
Offshore (long-term)	0.436	0.410	0.420	0.455	0.406	0.425

Table 6.3: *Simulated and calibrated CFs for on and offshore wind. The 'long-term' offshore CFs are calculated using all weights in Table 6.2 and the 'existing farms' only use those farms currently operational (highlighted in bold in Table 6.2).*

6.4 Validation

To check the validity of the wind modelling, a comparison between simulated and actual historic outputs was undertaken. Sadly data on individual wind farm output, or indeed high resolution (i.e, half-hourly) aggregated output was not available at the time this analysis was undertaken, although monthly resolution output data for GB can be extracted from OFGEM's ROC register [206]. Monthly CFs were derived from historic data on ROC certificate allocation for January 2005 - March 2009 by dividing the *'Number of certificates'* by *'Capacity'*. This data contains the ROC certificate data for over 700 wind farms throughout GB. Each wind farm was allocated to a region according to the polygonal regions in Fig. 6.1. Simulated monthly CFs could then be compared with both aggregate and regional CFs. Note that not all wind farms on the ROC register have been generating for the same length of time, therefore each month's aggregate CF includes only the sites that were in operation at that time.

A selection of onshore regional comparisons is provided in Appendix A.6 with the aggregate comparisons shown in Fig 6.2. The profile for aggregate onshore GB CFs shows a over-estimation in most periods. However applying the 0.919 scaling factor in Table 6.3 provides a better match.

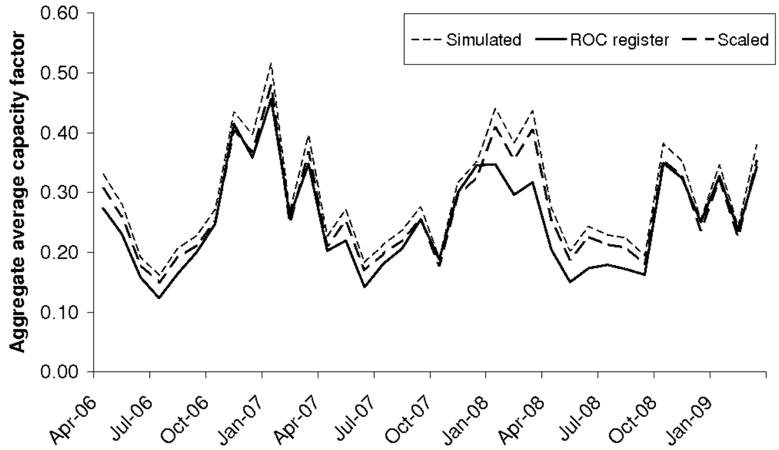


Figure 6.2: *Plot of comparison of simulated monthly aggregate GB onshore CFs (dashed lines) versus actual GB CFs estimated from ROC register data (sold line).*

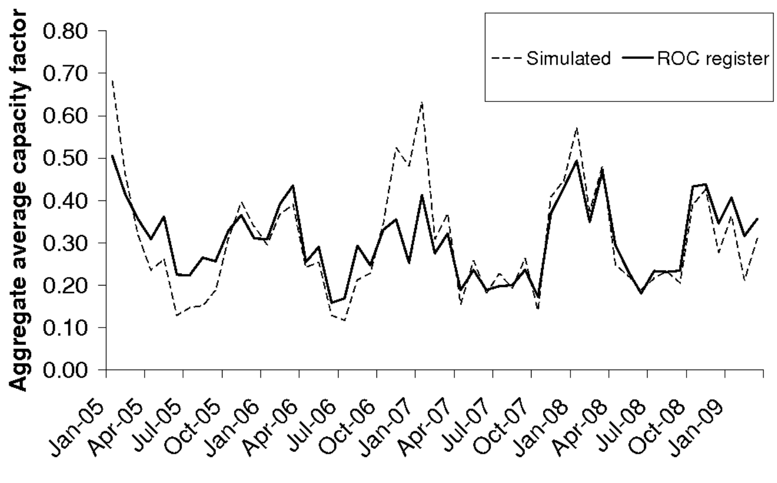


Figure 6.3: *Plot of comparison of simulated monthly aggregate GB offshore CFs (dashed line) versus actual GB CFs estimated from ROC register data (sold line).*

The aggregated and scaled offshore GB CFs are compared for all 13 wind farms in operation in Fig. 6.3. The graph shows a reasonable match in terms of pattern, although an over-prediction in winter and under-prediction in summer is witnessed in some years. This over-prediction can largely be attributed to technical availability. For instance, of the four operational offshore wind farms analysed (highlighted with a '*' in Table 6.2), the average technical availability for the winter of 2007 was 69.8% [205]. There is major uncertainty surrounding the technical availability of offshore wind, particularly during winter months when much of the North Sea is closed to shipping and so farm access may be restricted. Due to lack of data, this issue cannot be addressed further here, yet it is an important caveat for these results and must be considered when assessing levels of generation adequacy risk for particular investment scenario (cf. Section 7.5).

6.5 GB hourly wind production

Once the time series for for GB onshore and offshore CFs for the period 2005-09 have been obtained (43,824 load hours), the expected installed capacity must be allocated between onshore and offshore and the aggregated hourly GB wind production computed. For instance, given an assumed onshore wind capacity, w_{on}, and offshore capacity, w_{off}, the simulated hourly wind production (WP) in hour i is given by:

$$WP_i = GBCF_i^{on} \cdot w_{on} + GBCF_i^{off} \cdot w_{off}, \tag{6.3}$$

where $GBCF_i^{on}$ and $GBCF_i^{off}$ have been calculated using (6.1) for the onshore and offshore resource, respectively.

To begin, a base case wind capacity growth profile is constructed. Here total installed wind capacity is expected to increase linearly from 2010 levels up to 30 GW by 2020 with a maximum of 35 GW in 2025, after which it levels off, as shown in Fig. 6.5. The plot also shows the onshore and offshore penetrations. Onshore wind is increasing at a rate of 1 GW per year with a maximum of 13 GW while offshore is increasing at 1 GW per year until 2015 and then increasing to 2 GW after that. The allocation between onshore and offshore areas is consistent with the data in [61] and [197].

Fig. 6.4 shows how the relative levels of installed capacity within the fully aggregated GB capacity factors (6.1) increase throughout the simulation. This provides a visual representation of the 'contribution' from each region to the weighted CFs and overall GB wind production.

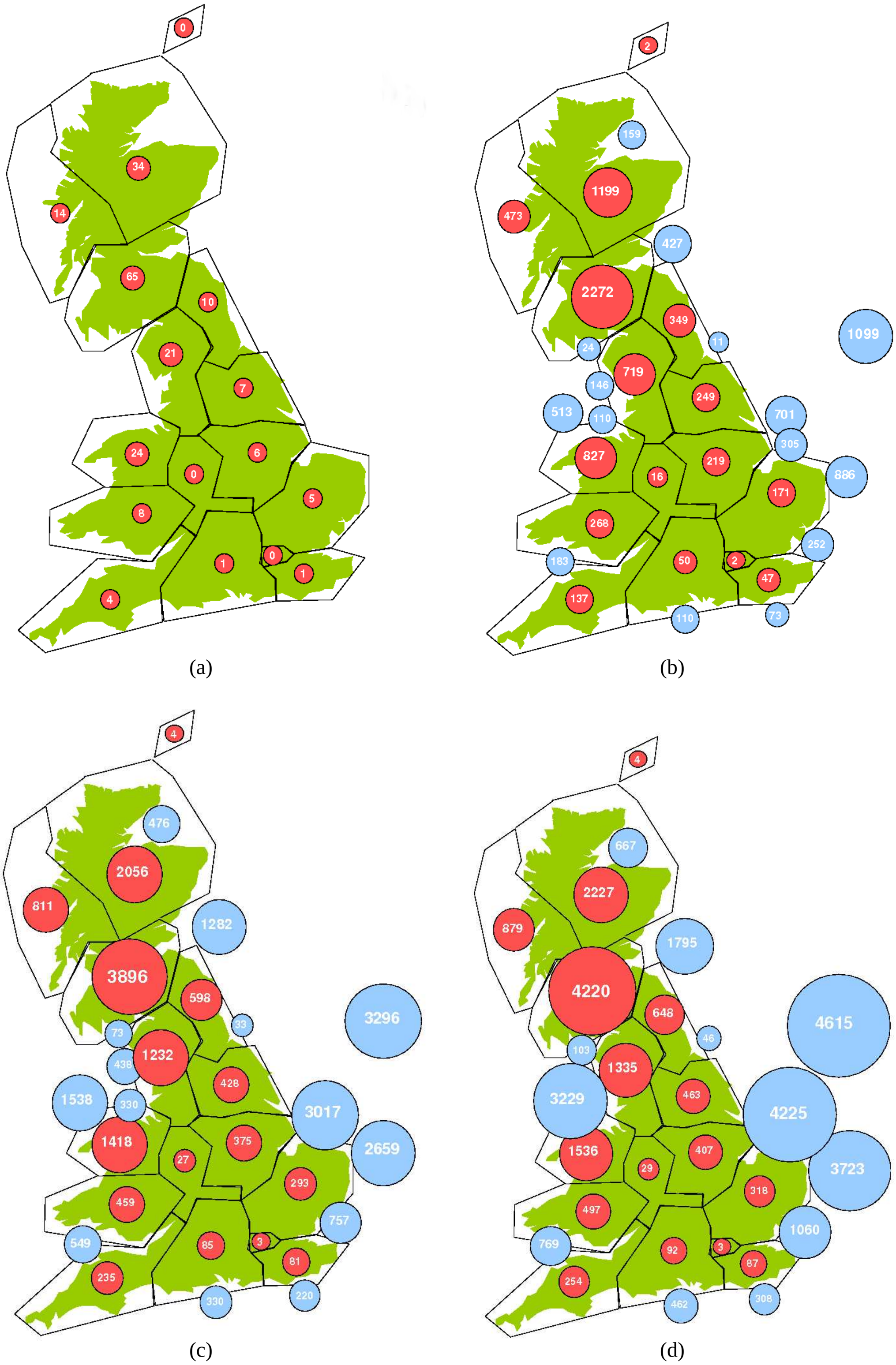


Figure 6.4: *Diagrams showing effective regional contributions to total wind capacity for (a) 2010 (no offshore wind); (b) 2015; (c) 2020; and (d) 2025.*

For instance, taking 2015 as an example, the total installed capacity of wind is 12 GW. This is split between 7 GW onshore and 5 GW offshore. Using (6.3), the simulated hourly onshore wind CFs (6.1) are multiplied by 7 GW to provide hourly onshore wind production, and the simulated hourly offshore wind CFs are multiplied by 5 GW to provide hourly offshore wind production. Thus, the absolute levels of installed capacity vary, but the simulated hourly CFs remain unchanged.

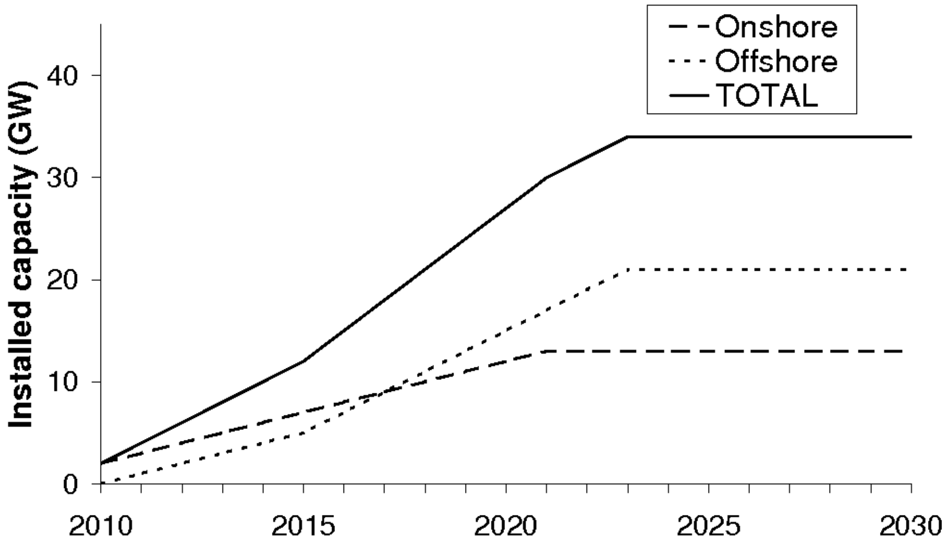


Figure 6.5: *Plot of total installed wind capacity growth assumptions. Distribution between onshore and offshore also shown.*

With this build schedule defined, an hourly residual load calculation using empirical load and simulated wind data can now be performed. The residual load in each hour is the hourly empirical load net of estimated hourly wind production (6.3). Fig. 6.6 provides a visualisation of the impact on the 2005-09 residual load histograms (considering 43,824 residual load hours for 2005-09) as penetration of wind increases. These histograms provide a graphical representation of the distribution of residual loads for a particular level of installed wind capacity. Note that the GB hourly loads for the period 2005-09 have been normalised by the year's average demand when constructing the histograms. A normalisation of the data is necessary when comparing multiple years to account for demand growth. These plots (penetration increasing in increments of 10 GW) do not match exactly the four diagrams in Fig. 6.4 (increasing in 5 year increments in line with Fig. 6.5), yet they demonstrate the change in residual load better than the equivalent 5 year incremental histograms.

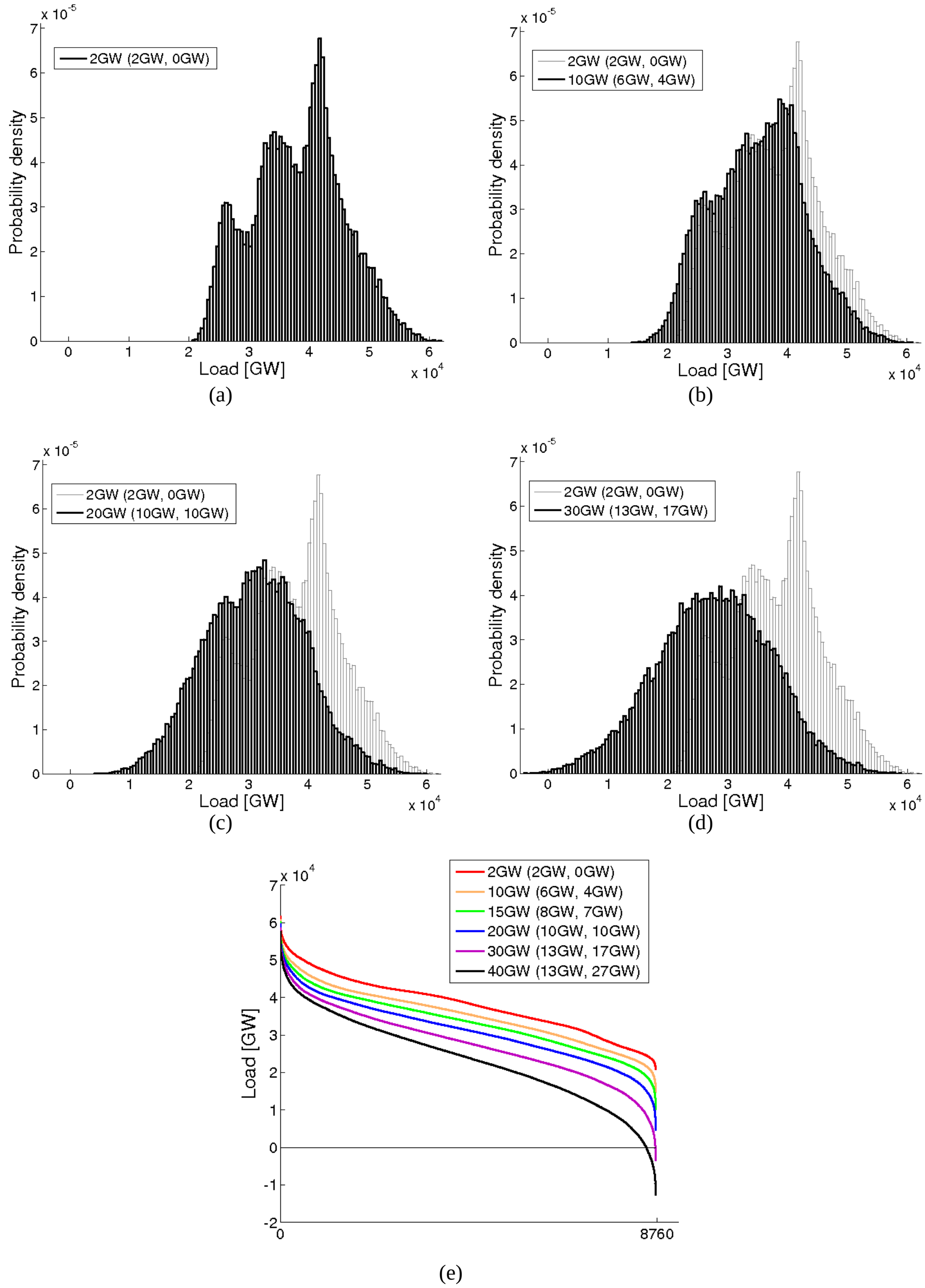


Figure 6.6: *Result of increasing installed wind capacity from (a) 2 GW to (b) 10 GW to (c) 20 GW to (d) 30 GW on residual load histograms. Also shown in (e) are some example residual LDCs. Numbers in brackets indicate volume of onshore and offshore capacity respectively.*

Interestingly, the residual load histograms become more 'Normal looking' as the penetration of wind increases. This phenomenon can perhaps be explained by considering the changes in the underlying data when moving from full to residual load at high wind penetrations. To illustrate this point, Fig 6.7 shows the normalised full load over time for January-March 2009 (blue). There is a distinct pattern to the full hourly load with daily peaks and troughs (a typical daily example is shown at the top of Fig. 6.7), although maximum and minimum daily loads change in relation to the season and day of week. These characteristics are reflected in the histogram in Fig. 6.6(a) by a high frequency of similar load levels (distinct peaks). By comparison hourly wind production is an independent random variable. So when deducting production from 30 GW from the full load at each hour, the profile is transformed into a seemingly random process with almost all of the distinct daily pattern removed. The Central Limit Theorem states that the distribution of the sum of a large number of independent random variables, each with finite mean and variance, will approach normality. This explains the migration from the distribution in Fig. 6.6(a) to the more "Normal" looking Fig. 6.6(d).

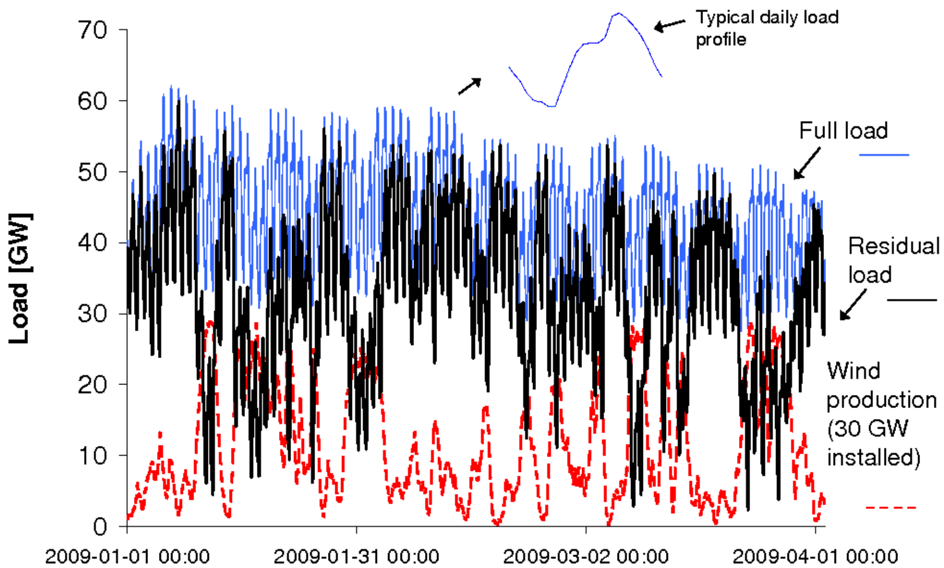


Figure 6.7: *Plot of normalised hourly full load, simulated wind production and residual load for 2009 with an installed wind capacity of 30 GW (13 GW onshore, 17 GW offshore).*

In addition, the standard deviation of the distributions increase substantially, with negative residual demands experienced at high penetrations. These observations are further highlighted in the residual LDC plots shown in Fig. 6.6(e); the negative loads occur for penetrations of

30 GW and above, though the peak load remains largely unchanged. In plain terms, these residual load histograms provide an approximation of what thermal generation production requirements (ignoring network issues) would have been for 2010-25 levels of installed wind under the assumption that simulated wind production and load for 2005-09 are representative of future patterns.

6.6 Capacity credit of wind

In this section the results of a capacity credit (CC) calculation for the simulated wind data is presented. CCs are an important addition to this wind analysis chapter and will be used in Section 7.5 to de-rate wind capacity at system peak in order to make projections about future levels of generation adequacy risk via the de-rated capacity margin.

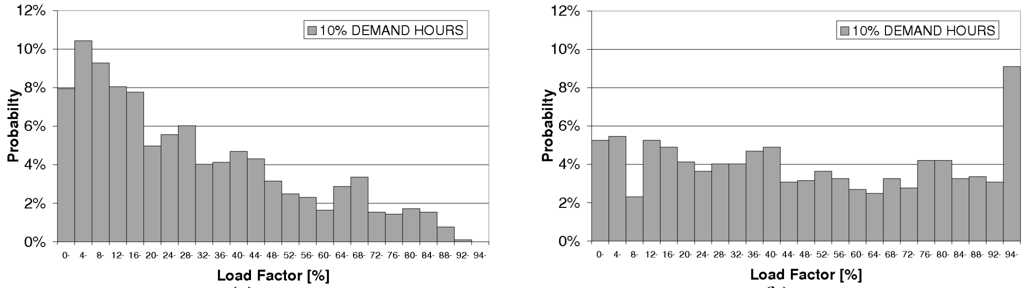
In this study, the "long-term" weighted GB CFs calculated above are used to calculate the ELCC, i.e., the amount of additional demand that can be supported after the wind is added whilst keeping the level of risk the same. In mathematical terms, this involves solving the following optimisation [93]:

$$p(X < d) = p(X < d - g^{+} + [ELCC]), \tag{6.4}$$

where X is available conventional generation (not including additional wind generation), d is demand and g^{+} is the wind generation. Following [61], the hourly simulated CFs for those hours within 10% of peak demand are used to represent the availability of the GB wind resource at peak demand (as in [61]). Over the period analysed (2005-09), this includes 1,045 demand hours. To illustrate, the probability mass function (pmf) for the weighted onshore and offshore wind CFs during these hours are plotted in Fig. 6.8. To compare, the pmf for entire 5 year period is given in Fig. 6.9. Plainly the shape, mean and standard deviation of the distributions are different in both cases, thus highlighting the difference in resource availability at peak compared with year-round. The plots also demonstrate the well known characteristic that the offshore wind resource is higher than onshore.

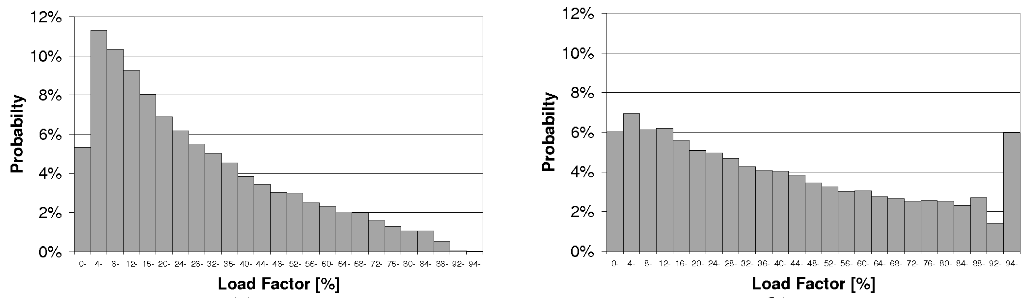
Using these probability distribution for wind availability, (6.4) becomes

$$p(X < d) = \sum_{j} p(CF = j)p(X < d - j \cdot W + [ELCC]), \tag{6.5}$$



(a) (b)

Figure 6.8: *Plot of probability mass function for available CF from GB (a) onshore (mean 0.307, SD 0.236) and (b) offshore (mean 0.493, SD 0.309) wind generation, based on simulated hours withing 10% of winter peak. If the CF falls in a particular range, it is deemed to be at the middle of that range (i.e. load capacity in the range 0-4% are deemed to be 2%.).*



(a) (b)

Figure 6.9: *Plot of probability mass function for available CF from GB (a) onshore (mean 0.291, SD 0.219) and (b) offshore (mean 0.416, SD 0.295) wind generation, based on all hours 2005-09.*

where $j = \{0.02, 0.06, \ldots 0.94\}$ is the midpoint of each CF "bin" (e.g., Fig. 6.8), W is the installed wind capacity (1-13 GW onshore, 1-22 GW offshore, cf. Section 6.5). For each j, the LOLP is calculated using the probability distribution for conventional generation, X, and demand level d (left hand side) or $d - j \cdot W + [ELCC]$ (right hand side). A fixed level of 60 GW for demand is assumed.

The results of the ELCC calculation, expressed as a percentage of installed wind capacity, are plotted in Fig. 6.10. The values range from 9-35% depending on level of installed capacity. The graph shows the typical pattern witnessed with wind CCs; as the penetration increases the CC value reduces. This demonstrates the dependence between sites due to "fuel" resource availability.

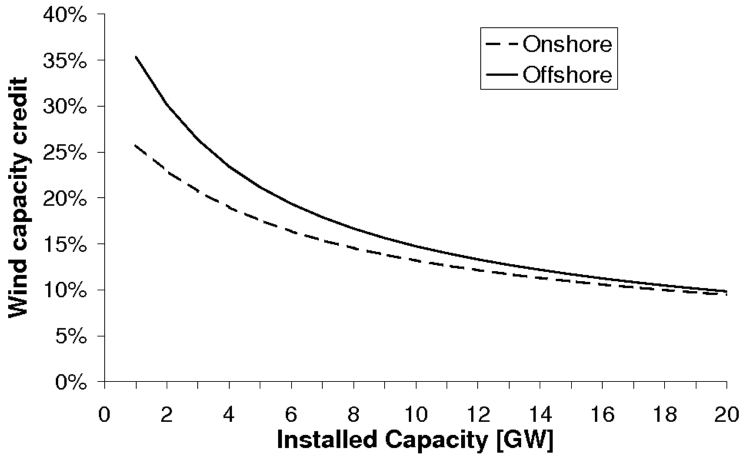


Figure 6.10: *Capacity credit results based on long-term weighted load factors; used in simulation model.*

6.6.1 Caveats

In this section, a specific methodology has been applied to calculate CC, however it should be noted this is not in itself cutting edge probability theory (e.g., see Dent [93]). For a superior approximation of CC in GB, including credible estimates for offshore wind, see Hawkins [204]. For instance, using the 5 year LOLE (i.e., sum of hourly LOLPs (2.12)), as in [204], would provide is more statistically robust estimate of the ELCC because it takes into account many more demand hours (e.g., around 18,000 winter demand hours between 2005-09) and thus provides a better representation of the relationship between demand and the wind resource availability at periods of peak demand. Further, in order to be confident about the contribution of wind generation to meeting peak demand, this would require an analysis based on more than 5 years of empirical demand and simulated wind data. For the purposes of the long-term dynamic simulation model presented here, these values provide a reasonable approximation.

6.7 Chapter summary

In this chapter the techniques used to estimate hourly wind production and resource reliability in GB have been presented. This began with the motives for developing a model that uses a hind-cast simulation in order to estimate the contribution of wind generation to serving load. This was followed by a description of the approach taken to model the GB onshore and offshore wind resource. Section 6.3 presented the wind model calibration against available published annual statistics, followed by validation against available wind generation output records. The results showed that simulated wind production provided a reasonable match with historic data, however mild scaling was required. Section 6.5 showed how for an assumed penetration of wind generation, hourly wind production and subsequent residual load is simulated. The shape of the residual load histograms changed significantly as the penetration of wind increased: they became more 'Normal looking', which was explained by the change from a distinct daily load over time profile to a more random profile due to wind variability. In addition, the standard deviation of the distributions increased substantially, with negative residual demands experienced at high penetrations. Finally, Section 6.6 explained how capacity credits for wind generation were estimated. The capacity credits results were 9-35% depending on level of installed capacity. The outputs from this stage of the work are taken forward into the next chapter where the dynamic investment model is updated in order to predict likely investment trends in GB with a high wind penetration.

Chapter 7
Improving the Investment Model

This chapter provides details of the changes made to the model logic on account of high penetrations of wind generation and increased uncertainty surrounding utilisation and prices. In summary, these changes were to 1) consider growth in wind capacity to be exogenous and 2) improve the method of calculating expected output, costs and revenue of thermal generation subject to varying load and random independent thermal outages, i.e., use a probabilistic calculation method for gross margins instead of a deterministic approach (5.16). In addition, a new probabilistic method for including price mark-ups due to market power in the gross margin calculation is included. Maintaining a GB focus, an application of the model to the GB investment market is provided, with particular attention paid to the evolution of generation adequacy risk over the next 30 years.

A sensitivity analyses on a number of key model assumptions provides insight into factors affecting the simulated timing and level of generation investment. This is achieved by considering the relative change in simulated levels of security of supply risk metric such as de-rated capacity margins and expected energy unserved. These results provide insights into the increased 'energy-only' market revenue risk facing thermal generating units, particularly peaking units that rely on a small number of high price periods in order to recover fixed costs and make an adequate return on investment.

7.1 Introduction

When making an economic assessment of the potential for generating capacity investments, there is a need to model varying loads (e.g., in the form of the LDC), the expected contribution of generating units to serving these loads, and the revenues they receive by doing so. It is helpful if the technique used is computationally fast, accurate and robust, especially when multi-year simulations of a market are to be repeatedly run. One approach is to use probabilistic production costing, a long established method for calculating the expected output and costs of a thermal generation system subject to varying load and random and independent forced outages [207, 208]. The first focus of this chapter is the integration of a probabilistic production costing method into the dynamic simulation model. The method considers the annual load curve and convolves it with generator outages using the Mix of Normals distribution (MOND) approximation. This production costing method was first described in [209] and then extended and used for the calculation of equilibrium capacity investment in a power market in [210] and [211], respectively. In this study, the method is applied for the first time to a nonequilibrium setting as part of the dynamic market simulation. Furthermore, the production costing methods used in previous dynamic models are deterministic, and therefore underestimate average costs (due to Jensen's inequality [212]). This can also be said for the calculation of gross margins used in the preliminary model implementation.

This chapter also calls on the wind modelling work of Chapter 6 to assess the impact of high penetrations of wind power on the investment risks associated with conventional thermal generation. Therefore the method above is extended to include results of the residual load calculation (load net of wind output) from Chapter 6. This residual load data is then used in the MOND production costing model in a new way. Finally, the MOND model is incorporated in the dynamic investment model and again applied to a simplified GB power system, this time for an assumed (exogenously increasing) installed wind capacity. Fig. 7.1 summarises this version of investment market model via a bull's eye diagram (see [115]) depicting endogenous, exogenous and excluded model parameters.

7.2 Updates to general investor logic assumptions

There are several adaptions to the model in terms of the investment logic. Investors are assumed to have the modelling capabilities available to formulate a reasonable approximation of the effect of wind generation on residual demand, thus the simulated hourly wind production data presented in Section 6.5 is employed. This maintains the adaptive expectations hypothesis design introduced in sub-section 5.3.1. Note that variations in weather patterns are likely to average out to zero over the economic lifetime of an investment and thus do not affect decisions about capacity expansion. Again, this makes the same assumption as in sub-section 6.5 that climate change does not impact on future weather.

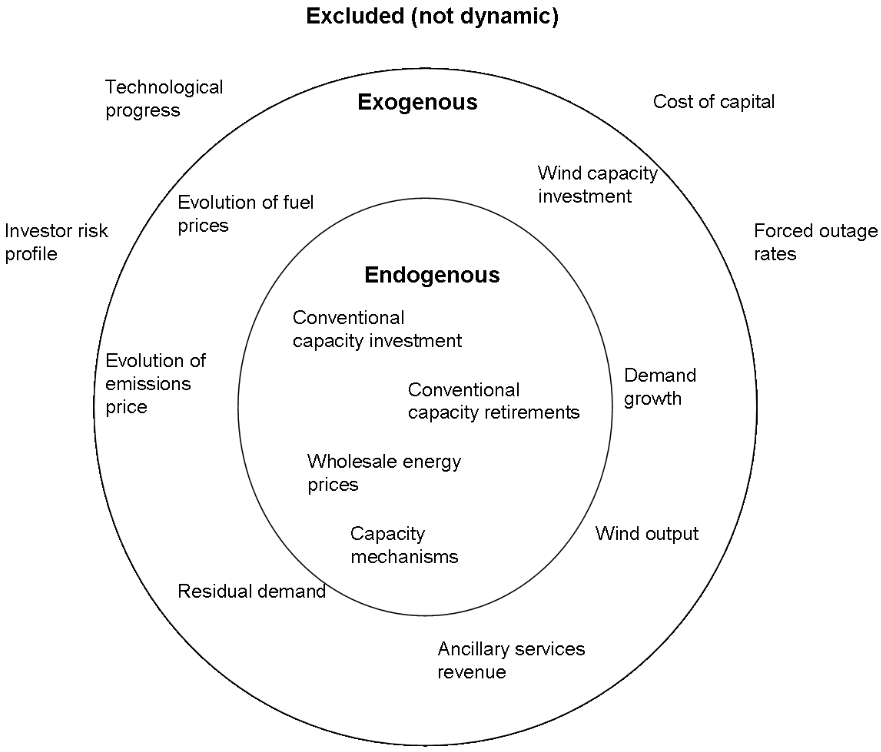


Figure 7.1: *Bull's-eye diagram of investment market model. Inspired by [115].*

In light of the recent government estimates, which were released after the initial model implementation, fuel and carbon prices are estimated using the DECC central case estimates [181, 213]. Assuming investor price forecasts are similar to these estimates,[1] the investor model uses the DECC estimate plus a random variable to estimate future fuel prices. This random variable is modelled as a mean reverting stochastic process with seasonality [136]:

$$dF_t = \alpha(MR(t) - F_t)dt + v(F_t)dW_t, \quad (7.1)$$

$$MR(t) = \frac{1}{\alpha}\frac{dq}{dt} + q(t), \quad (7.2)$$

where F_t is the fuel cost at time t, $q(t)$ is the DECC estimate (i.e., the 'seasonal' element), $MR(t)$ is the time dependent mean reverting level which depends on the DECC estimate, v is the volatility, λ is the speed of mean reversion and W is a standard one-dimensional Wiener process [32]. Initially, $\lambda = 0.7$ in all cases (indicating a reasonably short excursion length) and v is 5%, 7%, 10% and 20% for uranium, coal, gas and carbon prices, respectively to reflect

[1] Which in the case of natural gas match quite well with available future prices from ICE Futures Europe (out to 2017) but are arguably a little low for coal. At the time of writing, Newcastle futures were rising at about 1.5% annually not falling by 6% as suggested by DECC [214].

the relative price volatilities. An example of each random walk for 100 MC simulations for gas price is shown in Fig. 7.2. Note that although DECC provide estimates for future fuel and carbon prices, these should not be treated as a definitive answer; this would imply that DECC is able to fix markets, which is possible only for carbon, e.g., the UK Government recently announced a minimum floor price on carbon of 16 £/T in 2013, rising to 30 £/T by 2020 [215]. Note that, simlarly to the preliminary implmentation, no correlation between fuel prices was considered here.

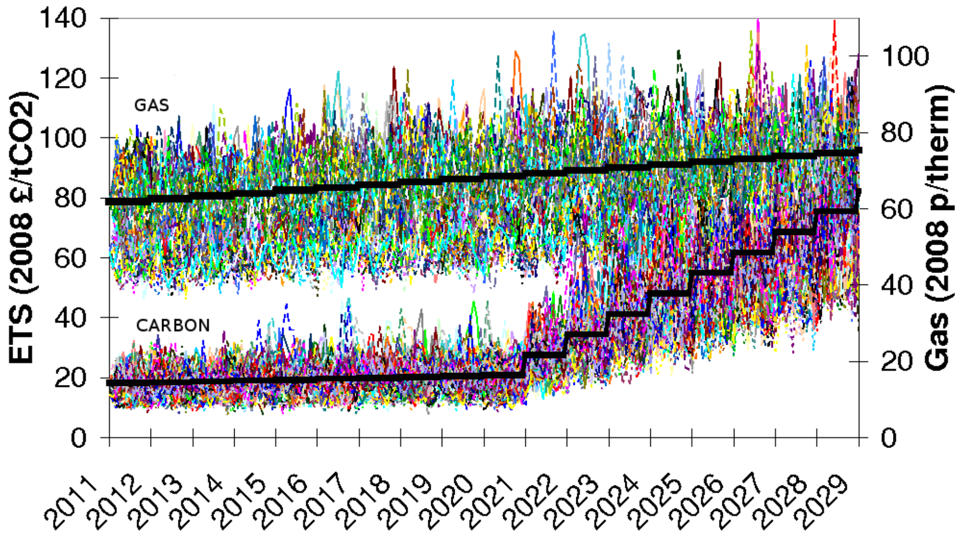


Figure 7.2: *Example of simulated random walks for gas (right y-axis) and ETS prices (left y-axis). DECC estimates are the solid thicker lines.*

Investors consider annual demand growth to be stochastic and this is sampled from a Normal distribution. Here a mean of 0% and standard deviation of 1% is assumed. This is based on variations in demand growth [54] as well as the perception that economic growth could be offset by increased energy efficiency, e.g., [68], thus allowing for small or even negative demand growth. This is consistent with recent government and GB System Operator (SO) projections about future electrical demand in GB [27, 216].

Only the first n years of expected revenues are stochastically simulated by the investor (here $n = 7$, previously $n = 15$); for the remaining years the (discounted) average of the simulated revenues are used (e.g., similar to [54, 65]). This assumes that simulated prices for the first 7 years of plant operation are representative of the total expected plant lifetime. Furthermore, investors cap the total expected annual revenues received from scarcity rents (sub-section 7.3.3) at the annualised cost of an OCGT. These actions ensure that expectations about future revenues are not unduly influenced by high forward simulated wholesale prices owing to generation retirements far out into the future. Note that no regulatory price caps are implemented in the real-time simulation. The decision to reduce the 15 year forward simulation time horizon ($n = 15$; sub-section 5.3.1) was also informed by the need to achieve a reasonable execution time, an issue that arises when moving from a 9 to a 30-year simulation time horizon.

To keep the model simple, revenue from STOR is simplified; here investors in peaking capacity (i.e., OCGT) assume an additional revenue of £10,000/unforced MW/yr can be obtained from the ancillary services (AS) market. As the results in sub-section 5.5.4 showed, AS revenues form a critical component of peaking capacity profitability, yet they are not sufficient by themselves to trigger investment and a combination of energy market and AS revenue is required in order to obtain adequate gross margins. Similar models applied in the US (e.g., [54]) also do not treat the AS market explicitly, and instead assume a fixed (and relatively modest) contribution of AS to peaking unit gross margins. Furthermore these are likely to be a second order effect when considering generation investments on a decadal time scale and are likely to be relatively unimportant for cycling and base load capacity, given their relatively large capital costs. Here investors will not include AS revenues in the profitability calculation if OCGT capacity exceeds 8 GW (i.e., volume of installed OCGT capacity at the start of the simulation, Table 7.5), which essentially limits the total obtainable revenue from AS if the volume of peaking capacity becomes large. On account of the point made above, this limit should not unduly dampen peaking capacity investment.

As already demonstrated in Section 6.5, residual load (load net of wind output) is likely to become more volatile as the penetration of wind increases (e.g., Fig. 6.7). A consequence of this is that more ancillary services will be needed. For instance, the GB System Operator (SO) recently forecast that operating reserve requirements will increase by 53% from 4,777 MW to 7,335 MW, between 2010/11 and 2020/21 [217] to compliment an increase in installed wind capacity from 5,795 MW to 24,599 MW over the same period (a growth pattern not too dissimilar from the one considered here). Securing adequate flexible capacity is addressed during operational timescales through the GB System Operator's reserve services (e.g., fast reserve, fast start, demand management and short-term operating reserve). The principle concern here is to determine whether sufficient capacity is built in order to meet periods of high demand and/or low available wind generation, so increases in these services and their impact on peaking capacity profitability are beyond the scope of this thesis.

Once again, because the model randomly samples capacity construction times, fuel prices and load growth, a large sample is required in order for investors to obtain reliable estimate of expected project value (5.17). Here, 100 MC simulation runs are carried out for each plant type in each decision year. If used for an actual policy analysis, the impact that the number of Monte Carlo runs has on the model outcomes should be evaluated and, if possible, larger sample sizes used. Here extensive testing of the effect of the number of simulations was not undertaken. An indication of the importance of sample size is given by the standard error of the expected revenues. The samples are independent and identically distributed (i.i.d), so this error is proportional to $1/\sqrt{N}$ (here $N = 100$) [218]. For instance, for the base case results presented in Section 7.5.1 for nuclear in 2020 in one run, the standard error was about 5% of the average revenue, which is reasonably precise. If the sample size was increased ten-fold to 1000 samples, one would expect an error of $5\%/\sqrt{10}$, or about 1.5%. Moreover, application of variance reduction methods (e.g., [219]) can result in smaller standard errors than i.i.d. sampling; however this is left for future research.

7.2.1 Mothballing

The decision to mothball (or de-mothball) is now taken annually and is based on the predicted gross margins over fixed operational costs (AGM) (£/MW) for the next three years of operation averaged over the MC simulation runs, i.e.,

$$AGM_x = \frac{1}{100} \sum_{j=1}^{100} \sum_{\forall j, i=1}^{3} \frac{GM_x^i - FC_x}{(1+r)^i}. \tag{7.3}$$

If this is negative (or positive) then some currently operational (or mothballed) plant will be mothballed (or de-mothballed). Note that no additional costs incurred as a result of mothballing or de-mothballing are considered here.

7.2.2 Modelling aggregate investment response

In some circumstances the expected profitability of new investments is extremely high, thus triggering a wave of new builds. Under these circumstances the investment rate will be limited by: 1) the firm's beliefs about how many other market participants will move to invest; 2) the impact of new investment on the profitability of their existing plant; and 3) on the ability of the firm to secure the debt required to fund multiple projects [121]. Using an aggregate investor response curve is useful in models of this type. For instance, in [54], the aggregate investment rate is increasing with the *"risk-adjusted forecast profit"*, which is derived from the investor's (risk averse, concave) utility function. Also in [121], an S-shaped nonlinear function of Profitability Index is used and various profitability functions are used in [127] to model investment rates based on managerial optimism concerning economic (i.e., expected profitability) and strategic (i.e., retaining market share) considerations.

In this model, a function is applied to the outcome of the VaR decision rule in order to estimate the aggregate investment response of the market. This function is increasing with the expected profitability and is given by:

$$\xi_i = max\left\{0, \xi_{max} \cdot \left(1 - e^{(-\beta \cdot PI_x)}\right)\right\}, \tag{7.4}$$

where PI_x is as in (5.21), ξ_{max} is the maximum yearly investment lump per technology x. ξ_{max} and shape parameter β are calibrated using the fixed assumptions that 1) zero investment is made if $PI_x^q < 0$ and 2) $\xi_i = i_x$ volume of investment is made if $PI_x = 1$, where i_x is a chosen fixed constant. Note that ξ_i provides the link to the state equation (5.3).

The function used in the base case is shown in Fig. 7.3 with fixed $i_x = 2$ GW and $\xi_{max} = 4$ GW, and $\beta = 0.7$ resulting from the calibration. Changing β alters the aggregate response, as shown by the dotted lines. There is the potential to include different response curves within each fuel or peak/base generator type (e.g., as in [121]). There is also an additional step whereby after each 2 GW of capacity of investment is triggered, the investor decision is re-run to ensure that no other plant types become more attractive in relation to other options as a result of this addition. This maintains the iterative adding characteristic described in sub-section 5.3.3. For multiple investments with $PI_x > 0$, the option with the highest PI_x is chosen. Finally, total annual investments are limited to 10% of total installed capacity.

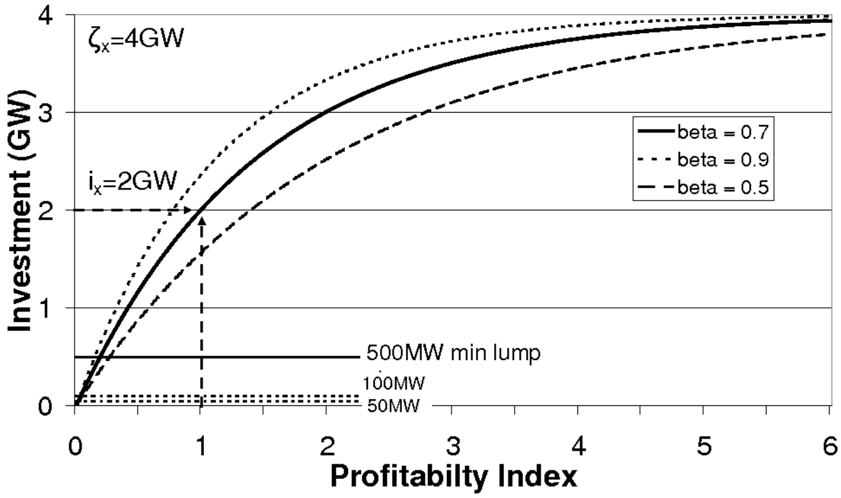


Figure 7.3: *Plot of model aggregate investment response curve defined by (7.4) (solid line) where $i_x = 2\ GW$ and $\xi_{max} = 4\ GW$. Also shown are the minimum investment lump sizes along with curves for different values of β.*

The degree of heterogeneous investment response is not modelled directly, however the investment response of the model is a smooth function of prices and costs, reflecting how a heterogeneous group of market participants would likely respond to changes in market conditions. In particular, the amount of investment is a smooth monotonic function of expected returns, as reflected in the AIRC (7.4), i.e., it is not a type of response where if profits exceed a threshold, a large amount of investment occurs.

Furthermore, forward and bilateral contracting between generators and suppliers is also not modelled explicitly, however both forward prices and investors' expectations are driven by the same market factors. Moreover those contracting are implicit in the way investor price expectations are modelled. In a commodity market, forward price expectations will be driven by the same economics as short-term prices. For instance, in the face of capacity retirement and demand growth, forward prices will rise. The investor simulates prices 7 years out, which is further ahead than most forward electricity markets go. Moreover explicitly representing contracts and forward price expectations would make the model significantly more complicated. For instance, [220] presents a single stage investment problem in the presence of endogenous contract and physical markets. Even in that very simple circumstance, the model and its analysis is very complex. The effect of more long-term forward and bilateral contracts would be to reduce revenue risk for investors, i.e., lowering risk aversion, thus making them more willing to invest for a given distribution of energy prices. Therefore the effect of more or less forward contracting can be implicitly considered by adjusting the risk aversion parameter of the VaR model (i.e., the level of risk aversion, q, or the WACC, r, in (5.17)).

The volume (GW) of plant mothballed or de-mothballed is determined simplistically using the linear function $\xi_i = min(M, AGM_x/10^4)$, i.e., decreasing or increasing with AGM_x (7.3) up to a maximum of M GW (Fig. 7.4). In this case M is chosen to be 2 GW. This was inspired by the literature review of Prospect Theory.

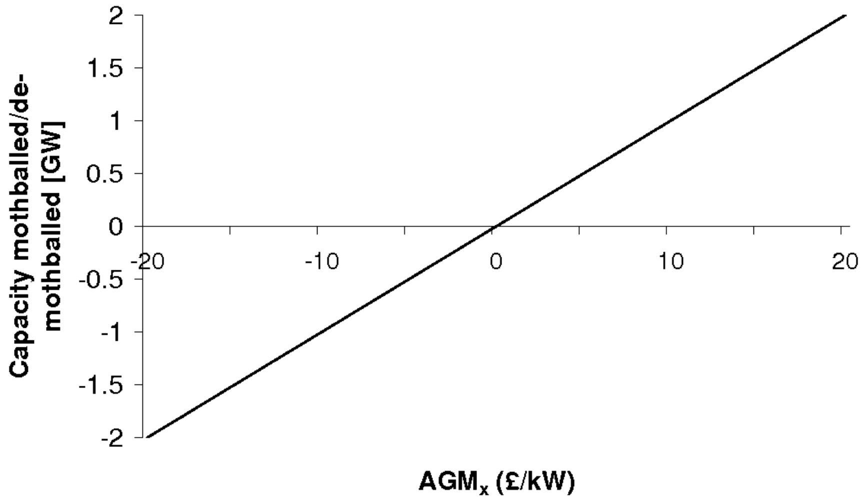


Figure 7.4: *Plot of model aggregate mothballing response curve. Note that the x-axis has been rescaled to (£/kW).*

7.3 Production costing by Mix of Normals

When estimating expected long-run production costs, there is a need for a reasonably accurate approximation of the LDC. In section 5.3.1, a 5^{th} order polynomial function was fitted to the simulated annual LDC. This approach, along with the method of calculating plant output was quite simplistic, and therefore it was decided to develop a more accurate method of approximation. A Mix of Normals distribution (MOND) approximation meets this requirement and is also very easy to convolve with plant outages for the expected output calculations. This motivated the decision to integrate this technique into the dynamic investment market model.

A MOND is described as follows: consider a set, $Y = \{y_1, \ldots, y_n\}$, of Normally distributed random variables with the i^{th} element having mean μ_i and variance σ_i. Let $\Phi(x|\mu_i, \sigma_i)$ be the cumulative distribution function (cdf) of y_i. A MOND is a convex combination of the Normal distributions and is defined by

$$F(x) = \sum_{i=1}^{n} p_i \Phi(x|\mu_i, \sigma_i), \tag{7.5}$$

with $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$, where p_i is the weight of the component y_i [209]. Let μ and σ be the mean and standard deviation respectively of the cdf described by (7.5). These parameters have the following properties:

$$\mu = \sum_{i=1}^{n} p_i \mu_i, \tag{7.6}$$

$$\sigma^2 = \sum_{i=1}^{n} p_i(\sigma_i^2 + \mu_i^2) - \mu^2. \tag{7.7}$$

Note that if X and Y are two independent random variables each with MOND $F_X(x)$ and $F_Y(y)$ given by (7.5), then $X+Y$ is a MOND. The proof of this property is given in appendix A of [209]. In this application the distribution of load (a MOND) is convolved with the distribution for available conventional thermal generation (cf. Section 7.3.1). It is a standard assumption in probabilistic costing and loss-of-load probability models that the outages of different generating units are independent of each other, and independent of load [64, 221]. The process starts by splitting the time horizon over which costs are calculated into periods. In this case, the duration of each period is one year; although shorter periods can also be used to account, e.g., for seasonal capacity. The expected load at each hour is a random variable.

A MOND fit for approximating the annual LDC (MW) is required. For example, if $f_L(x)$ is the probability density function of load and $F_L(x)$ is the cumulative distribution function of $f_L(x)$, then the LDC is simply the rotated and rescaled load *exceedence* distribution.[2] This is the inverse of $8760(1 - F_L(x))$ where

$$F_L(x) = \sum_{k=1}^{K} p_k \Phi_k(x|\mu_k, \sigma_k), \tag{7.8}$$

which is a mixture of K Normals (Φ_k) with the same properties as (7.5). For a particular K, the best fitting value of each μ_k, σ_k and p_k can be found by solving an optimisation problem

[2] The exceedence distribution gives $p(X > x)$, that is the probability that the random variable X (in this case load) is greater than x.

μ_1	μ_2	μ_3	μ_4
43802	41649	33971	26089
σ_1	σ_2	σ_3	σ_4
5743	1410	3372	1771
p_1	p_2	p_3	p_4
0.427	0.120	0.329	0.126

Table 7.1: *MOND fitted values for normalised 2005-2009 demand data, i.e., Fig. 6.6(a).*

that minimises the sum of squares of the difference between observed and fitted values of the LDC:

$$\begin{aligned} \underset{\mu_j,\sigma_j}{\text{minimize}} \quad & SS_{err} = \sum_{i=1}^{I} \left[L_e(q_i) - L_f(q_i)\right]^2 \\ \text{s.t.} \quad & L_f(x) = 1 - \sum_{k=1}^{K} p_i \cdot \Phi_k(x|\mu_k, \sigma_k) \\ & L_e(q_i) = \frac{2i}{2I-1} \\ & \sum_{k=1}^{K} p_k = 1 \\ & \mu_j \geq 0, \sigma_j \geq 0, p_k \geq 0. \end{aligned} \tag{7.9}$$

where the q_i are the rank ordered loads for the year, $L_e(q_i)$ is known as the *plotting position* of q_i (formula chosen to ensure $L_e(q_i) \neq 0$ or 1) [222]. In this application, $K = 4$ and in most cases $I = 8760$.

These results were computed using the MS Excel 2003 Solver add-in and for a mix of 4 Normal distributions $SS_{err} = 0.040$, compared with $SS_{err} = 11.499$ for a single Normal distribution. Plainly the objective function value for $K = 4$ is very good (at least a 10^2 improvement on the single normal case). As a result, a mix of 4 Normals is used.

To illustrate the accuracy of this technique at approximating a LDC, Fig. 7.5 shows the distribution of the GB hourly loads for the period 2005-09 (normalised by the year's average demand) and the fitted distribution. The distribution parameters displayed in Table 7.1 are used to fit the MOND. The improvement in fit when moving from one Normal distribution (Fig. 7.5(a)) to a mix of 4 Normal distributions (Fig. 7.5(e)) is clearly visible. In fact the difference between the two LDC plots (Fig. 7.5(f)) is not visible at this scale owing to the excellent fit provided by the MOND.

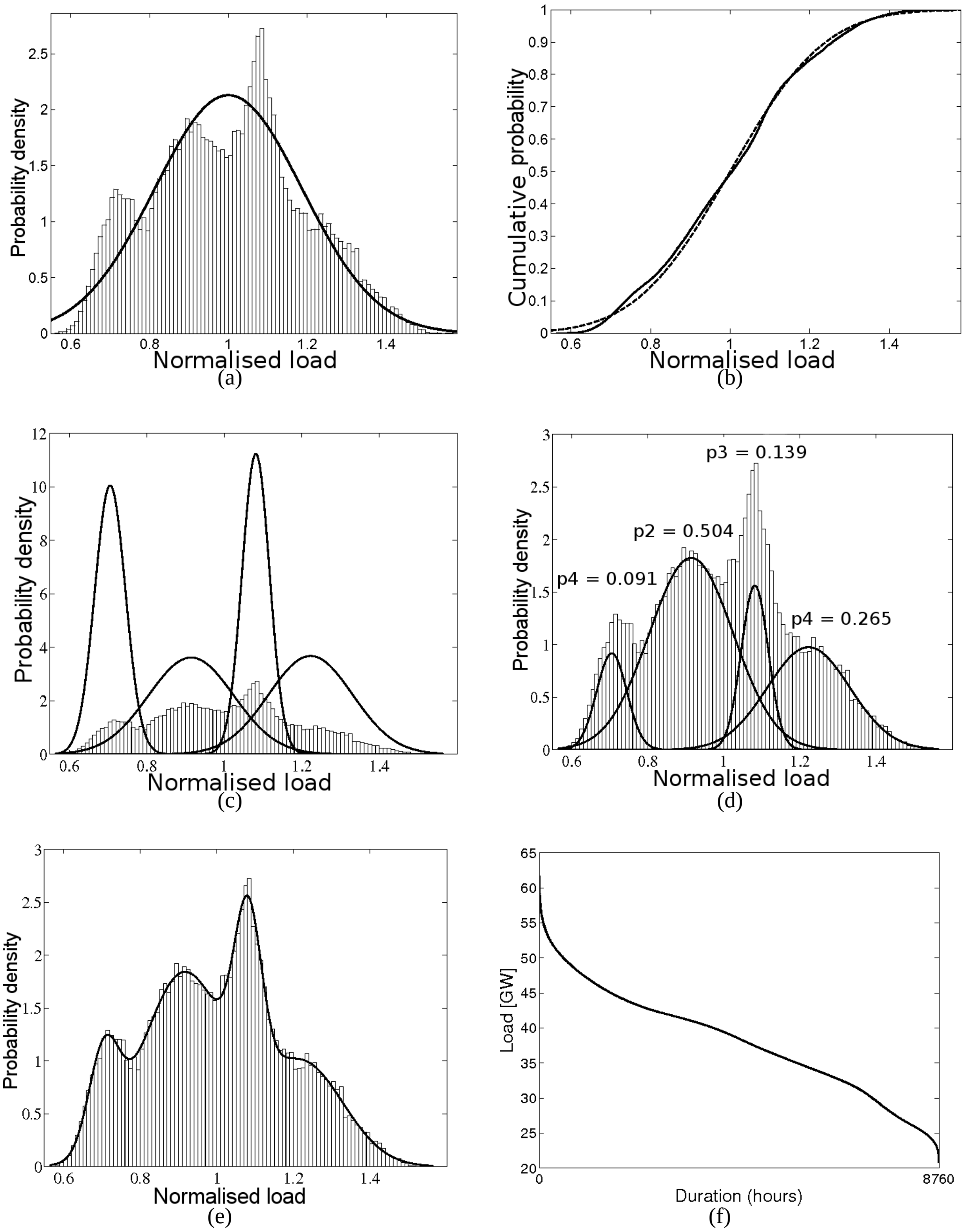


Figure 7.5: *(a) Histogram plot of normalised (by annual average demand) hourly load data and single Normal distribution fit, (b) Fitted cdf (dashed line) against normalised load data cdf (solid line), (c) Plot of each pdf for the 4 normals, 7.5(d) Pdfs scaled and labelled by p_i weights, (e) MOND pdf (4 Normal components clearly visible by their distinct peaks) and (f) LDC fit with negligible visible difference between mix of 4 Normals and empirical data.*

At this point it is worthwhile highlighting the key benefits of using a Mix of Normals distribution rather than the actual load duration curve. These are: 1) a MOND provides an accurate method of approximating load (see above); and 2) the concept of equivalent (or effective) load, that is the sum of the load (which is a MOND) and the outage capacity of conventional generators, can be approximated by performing a convolution of (Normal) distributions (see next sub-section). This concept of equivalent load plays a key role in probabilistic production costing [209]. Moreover this convolution technique offers computationally fast calculation of generator outputs, costs and revenues (see Section 7.5.1).

7.3.1 MOND with conventional thermal generation (after [209])

The next step is to estimate the available capacity for a set of generating units. Recall that the available capacity at each hour from a particular unit is a random variable which is characterised by the unit's forced outage rate (FOR). If m_u units of type n share the same capacity and FOR characteristics and are subject to independent forced outages, they can be treated as a single pseudo-unit (or generator) with a distribution with the moments (2.16) and (2.17) (subsection 2.8.4), where each (pseudo-) generator, G_n, has capacity $c_n = m_u c_u$. To simplify the presentation for the remainder of the chapter, the convention will be to use 'generator' when referring to a 'pseudo-generator' (collection of units of a given type), and c_n when referring to the capacity of that generator, because this is the last time individual units will be discussed.

Now the convolution property of the MOND is used to determine the distribution for the expected load still to be served after each generator is dispatched (in merit order). This distribution is called the *effective load duration curve* (ELDC) facing the next generator to be dispatched [221].[3] If the units can be grouped into N generators[4] each with characteristics (2.16) and (2.17),

$$L_n(x) = P\left\{L - \sum_{i=1}^{n} G_i > x\right\}$$

is defined as the load minus the available capacity of generator types $1 \ldots n$ ($1 \leq n \leq N$) [209]. If $F_n(x) = 1 - L_n(x)$ is the cumulative probability of effective load $x = L - \sum_{i=1}^{n} G_i$

[3] Note that Gross [209] uses the equivalent terminology: *equivalent load duration curve*.

[4] The more units in a group, the closer the Binomial distribution is to Normal.

facing the $(n+1)^{th}$ generator, then

$$F_n(x) = \int_0^{c_n} F_{n-1}(x+y) f_n(y) dy$$

and

$$L_n(x) = 1 - F_n(x), \tag{7.10}$$

where $f_n(y)$ is the pdf of the Normal that approximates the Binomial distribution describing the available capacity of the n^{th} type of generation with installed capacity c_n. Technically for a Normal distribution, the bounds in (7.10) should be $-\infty$ to ∞, but it is assumed here that the probability of falling outside the physically possible range $[0, c_n]$ is negligible owing to the Binomial distribution being rescaled to 0/c_u. $F_n(x)$ is computed by performing the convolution of the Normal distributions for load (which is a MOND) and available capacity.

The convolution property of a MOND [209] is applied, with generator loading carried out by merit order. The expected energy served e_n (in MWh/yr) by generator type n can then be given by

$$E[e_n] = 8760 \int_0^{\infty} \big[L_{n-1}(x) - L_n(x)\big] dx, \tag{7.11}$$

where $L_{n-1}(x)$ is the load still to be met after adding generator type $n-1$ and $L_n(x)$ is the load still to be met after adding generator n, which at the start of the convolution process ($n = 0$) is obtained from (7.8). The distribution for the ELDC after convolving in n generators is given by:

$$L_n(x) = 1 - \sum_{k=1}^{K} p_k \Phi_k(x|\mu_{L_k} - \sum_{i=1}^{n} \mu_{G_i}, \left[\sigma^2_{L_k} + \sum_{i=1}^{n} \sigma^2_{G_i}\right]^{\frac{1}{2}}). \tag{7.12}$$

Thus, the ELDC is described by a MOND with the same number of component Normals as the original LDC. An example of the iterative convolution process is shown in Fig. 7.6. The diagram depicts how the remaining expected load to be served is reduced each time a group of generators is convolved with the ELDC.

Given each $L_n(x)$ ($n = 1 \dots N$), the probability that generator n or higher (in the merit order) is the marginal source of energy (and so sets the price) is given by $L_n(0)$, i.e., the point where the function crosses the vertical axis in Fig. 7.6. Further,

$$h_n = L_{n-1}(0) - L_n(0) \tag{7.13}$$

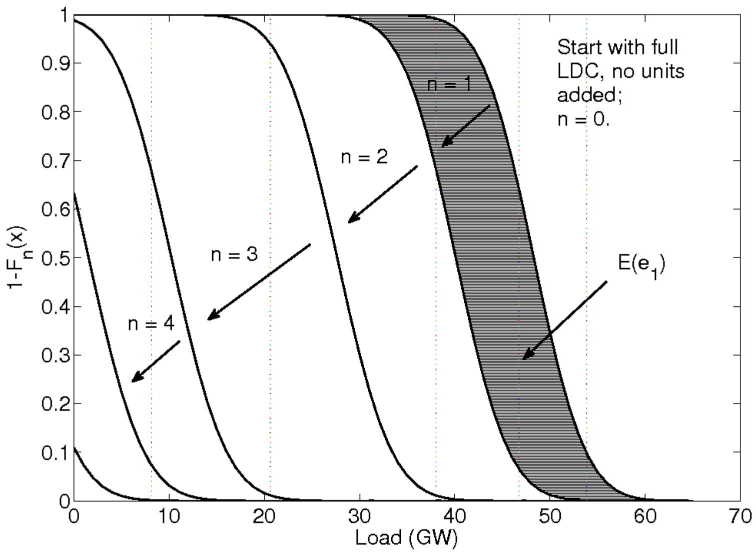


Figure 7.6: *Example of the convolution process. Shaded region is the expected energy served by the first unit dispatched (i.e., the result of (7.11) with $n = 1$).*

is the probability that generator n is on the margin. This result is used in sub-section 7.3.3 to calculate expected revenues per MW for a generator belonging to this generator type in an 'energy-only' market setting.

The unserved energy (MW) is then simply any positive remaining load once all N generators have been added. The annual expected energy unserved (EEU) per year can be calculated by

$$EEU = 8760 \int_0^\infty L_{N+1}(x)dx, \tag{7.14}$$

and the Loss-of-Load Expectation (LOLE) (hrs/yr) for the period is determined by [64]

$$8760 \cdot L_{N+1}(0). \tag{7.15}$$

7.3.2 MOND with a thermal-wind system

The interdependency between wind plants means that the convolution technique used to model thermal generation in [209] cannot be directly used to account for wind power generation. To address this issue, it is possible to construct an LDC that is net of wind production. An exogenous wind capacity is assumed and the resulting residual LDC facing thermal generation is computed and the MOND approximation is applied to this wind-adjusted data set. The residual load is simply the load minus the output from all wind plants at each hour, as described in sub-section 6.5. By taking this approach, the analysis can take into account both spatial correlations and seasonal (e.g., monthly and diurnal) trends in wind availability and their relationship with demand. This is believed to be the first time the MOND production costing technique has been applied using credible estimates in time and space for production from the combined on and offshore GB wind resource.

Furthermore the residual load approach allows us to calculate the number of hours that available wind generation is greater than aggregated demand, i.e., those hours when wind sets the system marginal price. This is computed using the residual load exceedence distribution before convolving any of the available thermal generation, i.e., the inverse of $8760(1 - F_L(0))$ where $F_L(x))$ is given by (7.8). This is of particular interest in systems with high penetrations of wind when direct wind dispatch may be required to curtail production at times when available wind exceeds demand.

Wind production may need to be curtailed under other circumstances. In particular, this implementation of the MOND technique does not consider the possibility of available wind generation exceeding either 1) available export capacity in a generation pocket due to transmission congestion or 2) raw demand net of inflexible base load (e.g., nuclear). Nor can the load duration curve method (of which the MOND is a particular case) consider the possibility that ramp rate limitations could also result in wind spill. Considering each in turn, given that the model is single bus, number 1) cannot be addressed here and is left for future research.[5] In this application, it is assumed that UK Government policy will ensure that the amount of congestion in the future will not be so large as to affect the basic economics of thermal generation investment. This is consistent with the GB regulator's 'connect and manage' transmission access policy [223]. In the case of number 2), if inflexible base load generation needs to be kept running then this will affect the economics of the wind generator being constrained (i.e., the wind generator is given a congestion payment by the System Operator). This is not considered in this analysis as installed wind capacity is an exogenous model parameter. The presence of inflexible generation can also affect the economics of baseload generation by increasing the number of hours of zero or negative prices. However the amount of hours when this occurs is small relative to

[5] If a representation of the load and available wind production in each region of the network is available, then curtailment of wind production due to transmission congestion could in theory be assessed by multi-area production costing methods.

the total number of running hours across the lifetime of the plant. For instance, for the GB case study presented in Sections 7.4-7.5, under the assumption that all plants can be turned down except nuclear, in no year and in no scenario is the probability of the net demand being below the expected available nuclear capacity greater than 7%. Future applications of this method could approximate the effect of inflexible generation by dispatching its must run capacity first and assuming a zero or negative price for the portion of the time that this capacity is on the margin.

It is important to note that the residual LDC approach removes the chronological issues that arise in the wind and load time series. This is particularly important in the presence of large amounts of hydro and pumped hydro generation where chronological production costing methods may be preferred to load duration curve methods. However this implementation of the MOND technique is applied to the GB power system where the amount of hydro and pumped hydro is relatively small (about 4% of capacity), so the use of a load duration curve approach is reasonable.[6]

7.3.3 Expected revenues from the energy market

In the absence of market power and unless there is a capacity shortage, the market price for energy will equal the marginal cost of the last generator to be dispatched. If there is a capacity shortage, and assuming there is no price cap, the price will clear at the level necessary to ration demand to the available capacity. More precisely, because consumers are not generally exposed to the real-time price, there is a willingness of suppliers to pay up to the VOLL when there is a shortage. In a more general case, some or all customers are exposed to the real-time price. This methodology is applied in [210] in a long-run equilibrium model; that approach is extended here to a dynamic long run nonequilibrium setting.

During a particular year, the probability that generator n will be at the margin is given by (7.13) and the price in that event will be the SRMC of the generator, $SRMC_n$, assuming price-taking (competitive) behaviour. Furthermore, once the convolution process has been completed for all N generators, the probability that there will be insufficient generation to meet demand is given

[6]Note that there is little scope for new build of hydro technologies in GB due to the lack of suitable sites.

by $L_{N+1}(0)$ and under this condition the price is assumed to reach VOLL. Using the result in (7.13), the expected gross margin for a particular MW of capacity belonging to generator n when generator i is at the margin is given by (£/MWh):

$$R_n^i = max\left\{h_i(\pi_i - SRMC_n), 0\right\} \tag{7.16}$$

where π_i is the wholesale price when generator i is at the margin, which in the absence of market power is given by $SRMC_i$. Using (7.13) to calculate (7.16), the expected annual perfectly competitive gross margin used in (5.17) can be calculated by (£/MW/yr):

$$CGM_n = 8760(1 - \rho_n)\Big[\sum_{i=1}^{N} R_n^i + (h_{N+1}(VOLL - SRMC_n))\Big]. \tag{7.17}$$

These are precisely the scarcity rents described in Chapter 2, which in a simple long-run competitive 'energy-only' market model, are just high enough to cover fixed costs and trigger investment [3]. The main numerical work here is in computing the $L_n(x)$ in (7.11); once these are known, the revenues are calculated easily by multiplying by the price - marginal cost differentials. Recall that ρ_n in (7.17) is generator n's FOR, hence the impact of generator outages on their expected gross margins is properly accounted for. Once again, no start-up or no-load costs are considered here.

Here the new probabilistic method for calculating revenues from price mark-up us presented. Recall that the use of the price mark-up function (5.10) was justified based on comparisons of the GB forward market index price (sub-section 5.1.4) and simulated prices. By comparison in this example, to keep the derivations simple, only the exponential function $w_1(L, G_N^*) = ae^{b\cdot(L - G_N^*)}$ (5.8) is used. Fig. 7.7 shows an example of the price function given by (5.11) when $w(L, G_N^*)$ is modelled using the exponential (5.8); the curve behaves like a classical linear step-wise marginal cost supply function for small loads, but as the system approaches scarcity, the mark-up function becomes evident and soon dominates the pricing mechanism.

Now the market price (5.11) no longer just depends on which generator is on the margin, which is the case under the classic perfectly competitive market case (7.17); it now depends on the overall margin, $G_N^* - L$ as well. Furthermore, since the price can exceed $SRMC_n$ if n is on the margin, the question of whether a particular incremental MW of capacity within n is called upon or not must be considered. This is because marginal generator n can still earn a positive gross margin. Thus, when generator n is on the margin, (5.11) must be calculated considering

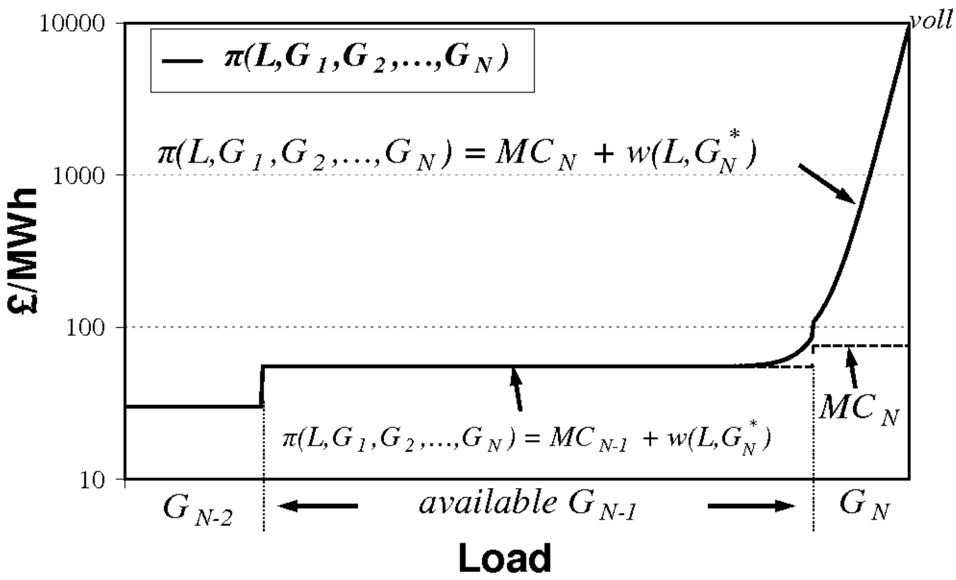


Figure 7.7: *Supply function for a given realised available capacity for generators $N-2$, $N-1$ and N with mark-up function defined by $w(L, G_N) = ae^{b(L-G_N^*)}$ (shown as black line). Marginal cost dashed, price is solid line.*

the probability that a particular MW belonging to generator n is dispatched or not. This is important for calculating expected returns on investments.

If R_n^{L,G_N^*} is the gross margin received by a MW of capacity from generator n for load L and total available generation G_N^*, then R_n^{L,G_N^*} will be calculated by one of the following means:

1. If the marginal generator, say i, is below n in the merit order (i.e., has a lower marginal cost) then the generator n will not be dispatched and $R_n^{L,G_N^*} = 0$.[7]

2. Else if the marginal generator has a higher marginal cost than n, then the probability of dispatch is 1 and $R_n^{L,G_N^*} = mc(L, G_1, G_2, \ldots, G_N) + w(L, G_N^*) - SRMC_n$.

3. Else if u *is* the marginal generator, then the probability of dispatch is between 0 and 1 and

$$\begin{aligned} R_n^{L,G_N^*} &= mc(L, G_1, G_2, \ldots, G_N) + w(L, G_N^*)p_n^{disp} - SRMC_n \\ &= w(L, G_N^*)p_n^{disp}, \end{aligned} \tag{7.18}$$

[7] Thus it is assumed for simplicity that dispatch is still in merit order, despite the presence of market power. However in oligopolies with asymmetric generating companies, a small high cost generator might produce power before a large low cost generator, as the latter is more likely to withhold capacity. This possibility is not considered here.

where p_n^{disp} is the probability that a MW belonging to generator n is dispatched.

This probability of dispatch is approximated by

$$p_n^{disp} = \max\left(\min\left\{1, \frac{-M_{n-1}}{c_n(1-\rho_n)}\right\}, 0\right), \tag{7.19}$$

where $M_{n-1} = \sum_{i=1}^{n-1} G_i - L$ is the surplus margin after L has been met using all available generation lower in the merit order than n (e.g., Fig. 7.8). In plain terms, (7.19) is the ratio of the MWs of generator n that are dispatched to the average total MW of n that are available (which differs from the actual available MWs of n). In the event that this margin is negative, the fractional term in (7.19) will be greater than 1, therefore the min $\{\_\}$ function is required in order to eliminate this possibility. This method of calculating dispatch probabilities is required to account for market price mark-up in (5.11).

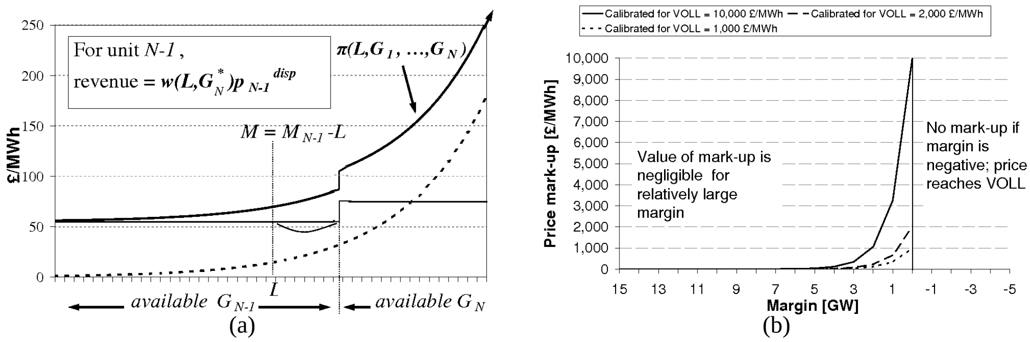


Figure 7.8: *(a) Aggregate supply curve showing price (solid upper line) for load L and revenue for generator of type $N-1$. Mark-up function also shown (dashed line). (b) Shows price mark-up for different values of capacity margin and calibrations for $a = 10,000$, $2,000$ and $1,000$.*

The expected gross margin (7.17) must now be extended to consider the price function (5.11) and merit order operation. This is less straight-forward than under marginal cost-based pricing (7.16) because the price mark-up requires consideration of the total (generation-load) margin, M_N, as well as the marginal unit. By assuming that price mark-up is non-zero (i.e., $w(L, G_N^*) > 0$) only when generator N or $N-1$ is on the margin, calculation of the probability distribution of $w(L, G_N^*)$ can be achieved by considering the joint probability distribution of just the capacity margins M_{N-1} and M_N. This assumption is reasonable in an aggregated capacity model where the generator size is large. As already discussed, empirical evidence from the GB market (e.g., [98]) is that mark-up tends to occur predominately during peak periods when surplus margins are relatively small, and where the surplus of available resource in other periods means the presence of market power is unlikely. Therefore, it is reasonable to assume that price mark-up is significant only when peaking plants N or $N-1$ are on the margin.

7.3.4 Expected price mark-up calculation

Firstly, for each component of the MOND (7.5), the joint distribution of capacity margins M_{N-1} and M_N, given by $f(M_{N-1}, M_N)$ is considered. The correlation is calculated as:

$$\begin{aligned} corr(M_N, M_{N-1}) &= \frac{cov(M_N, M_{N-1})}{\sigma_{M_{N-1}}\sigma_{M_N}} \\ &= \frac{\sigma^2_{M_{N-1}}}{\sigma_{M_{N-1}} \cdot \sigma_{M_N}} \\ &= \frac{\sigma_{M_{N-1}}}{\sigma_{M_N}}. \end{aligned} \tag{7.20}$$

The proof of this property is standard and provided in Appendix A.7.

Fig. 7.9 shows a plot of the *isoquants* of the bivariate Normal density for $\{M_{N-1}, M_N\}$, which are highly positively correlated owing to increases in M_{N-1} increasing M_N; these show the combinations of M_{N-1} and M_N that yield the same price mark-up. For a particular point (M_{N-1}, M_N), the diagram shows the value of the joint pdf (ellipses) and price mark-up isoquants (dotted lines in each quadrant centred $(0, 0)$). By splitting the space into four quadrants and sketching the isoquant maps, the effect of available capacity margins on plant revenues can be assessed. For instance, when plant type N is not dispatched (north-east quadrant), increases in M_N result in a decrease in mark-up (indicated by parallel horizontal lines, three example mark-ups shown). Likewise, when N is dispatched (north-west quadrant, M_{N-1} is in shortage). If M_N is in shortage (south-west quadrant), the mark-up is zero for all combinations of M_N and M_{N-1} (i.e., system is short of resource and the price will reach the VOLL). Note that the south-east quadrant is assumed to have zero probability; it would be calculated as having a nonzero probability only because the Normal approximation will give an unrealistic nonzero probability for negative availability capacity for generator $N-1$. In the application, this probability is negligible.

The ellipses in Fig. 7.9 are centred at $(E(M_{N-1}), E(M_N) = E(M_{N-1}) + (1-\rho_N) \cdot c_N)$. On the x-axis is the surplus margin after the load has been met using all available generation lower in the merit order than N, namely M_{N-1} (the expectation of which is 10 GW). On the y-axis is (for a corresponding $MN - 1$ on the x-axis) the expected surplus margin after the load has been met using all available generation, namely M_N (the expectation of which is 15 GW). So in this example, given that $E(M_{N-1})$ is positive, the expectation here is that the load can be met without generator N being dispatched. Furthermore, the difference $E(M_N) - E(M_{N-1}) = (1 - \rho_N) \cdot c_N)$ (5 GW) gives the expected available capacity of generator N.

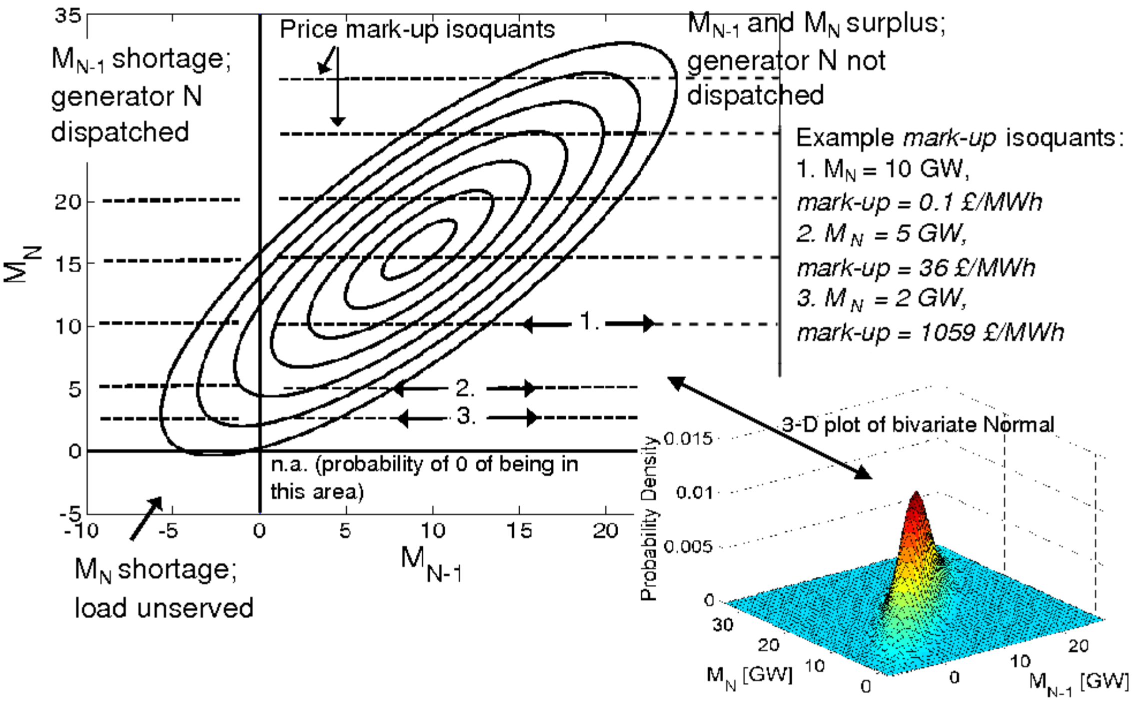


Figure 7.9: *3-D plot of bivariate Normal of $\{M_{N-1}, M_N\}$ (right) and its image on a 2-D plane (left) with isoquant maps for the price mark-up element of (5.11).*

Next, for each component of the MOND (7.5), the revenue that a MW belonging to generator $n \leq N$ earns from price mark-up, RPM_n (£/MWh), is considered for the following cases:

1. If $n < N - 1$ (i.e., n is lower in merit order than $N - 1$), then the expected revenue from price mark-up for a particular MW of that capacity is given by

$$RPM_n = \int_0^v (1 - \rho_n) f(M_N) w(M_N) dM_N, \tag{7.21}$$

where $f(M_N)$ is the pdf of the surplus margin $M_N = G_N^* - L$. This assumes that $w(M_N)$ is a function of M_N only, i.e., the overall system margin, and that generators $N - 2$ and $N - 3$, etc are all fully dispatched when mark-up is nonzero. The integral lower bound is zero due to price mark-up being zero if $M_N < 0$. More precisely, if the overall margin is negative then there is no mark-up and the price is set by the marginal cost of demand (i.e., VOLL). The integral upper bound in (7.21) is some value, v, above which the price mark-up is negligible owing to the large surplus margin (e.g., 7 GW in Fig. 7.8(b)).

2. Else if $n = N-1$, then RPM_n is broken down into two sub-cases:

   (a) If $M_{N-1} \leq 0$ (i.e., north- or south-west quadrant of Fig. 7.9) then

$$RPM_u = \int_{-\infty}^{0} \int_{0}^{v} (1-\rho_n) f(M_N, M_{N-1}) w(M_N) dM_N dM_{N-1}, \tag{7.22}$$

   which is the case where all of generator $N-1$'s available capacity will be dispatched because load exceeds the available capacity of generators 1 through $N-1$. Again, the inner integral lower bound is 0 because $M_N \leq 0$ results in zero price mark-up (south-west quadrant of Fig. 7.9).

   (b) Else if $M_{N-1} > 0$ (which implies $M_N > 0$, so north-east quadrant of Fig. 7.9) then

$$RPM_u = \int_{0}^{v} \int_{M_{N-1}}^{v} p_{N-1}^{disp} f(M_N, M_{N-1}) w(M_N) dM_N dM_{N-1}, \tag{7.23}$$

   where p_{N-1}^{disp} is the probability of dispatch of generator $N-1$, given by:

$$p_{N-1}^{disp} = \frac{(1-\rho_{N-1})M_{N-2}}{-1 \cdot (M_{N-1} - M_{N-2})} \tag{7.24}$$

   where M_{N-2} is the surplus margin after all available generation lower in the merit order than $N-1$ has been dispatched.[8] M_{N-2} is a random variable and computation of its pdf is awkward; however it can be approximated as follows: $M_{N-2} \approx (E(G_{N-1}) - M_{N-1})$. This is achieved by approximating the realised value of G_{N-1} by its expectation: $E(G_{N-1}) = c_{N-1}(1-\rho_{N-1})$.[9] Leading to:

$$p_{N-1}^{disp} \approx \frac{(1-\rho_{N-1})(c_{N-1}(1-\rho_{N-1}) - M_{N-1})}{-1 \cdot (M_{N-1} - (c_{N-1}(1-\rho_{N-1}) - M_{N-1}))}, \tag{7.25}$$

[8]The -1 scalar is applied to the denominator of (7.24) on account of M_{N-2} being negative. If it was positive, then G_{N-1} would not be dispatched (i.e., $p_{N-1}^{disp} = 0$), which is not considered in (7.23).

[9]Note here that c_{N-1} is the capacity of the generator type $N-1$, which is the sum of a number of individual units who share the same capacity and FOR characteristics.

i.e., the expected surplus margin over generator n's expected available capacity. Note that $M_{N-1} > 0$, so $M_{N-2} < 0$,[10] and thus $0 \leq p_{N-1}^{disp} \leq 1$.

3. Else $n = N$ and $M_{N-1} < 0$ and $M_N > 0$ (i.e., north-west quadrant of Fig. 7.9):[11]

$$RPM_n = \int_{-\infty}^{0} \int_{0}^{v} p_N^{disp} f(M_N, M_{N-1}) w(M_N) dM_N dM_{N-1}, \tag{7.26}$$

where

$$p_N^{disp} = \frac{(1-\rho_N)M_{N-1}}{-1 \cdot (M_N - M_{N-1})}, \tag{7.27}$$

i.e., the case where M_{N-1} is in shortage and some volume of capacity from generator n ($= N$) will be required to meet L.[12]

Finally, by integrating over the subregions of the $\{M_{N-1}, M_N\}$ space in Fig. 7.9 (i.e., using one of the cases 1-3 above), the expected annual gross margin, which is used in (5.17), for generator n can be calculated as

$$GM_n = CGM_n + E[e_n] \cdot RPM_n. \tag{7.28}$$

This is repeated for each component of the MOND for each decision year (suffix i in (5.17)).

This is an important extension to calculating the competitive gross margin (7.17); by exploiting the properties of the probability distribution of capacity margins, this allows for the anticipated additional revenue received from market price mark-up to be calculated during the production costing process. Procedures for calculating price mark-ups in probabilistic production costing models have been proposed (e.g., see [224]) but this is believed to be the first time the derivations (7.21) - (7.27) have been presented.

The MOND approximation technique is embedded within the dynamic investment model depicted in Fig. 5.1. More precisely, it is used to calculate the expected gross margin, GM_x^i, in

[10] In general, M_{N-2} could be positive, however the assumption here is that it is negative when price mark-up is greater than zero.

[11] There is no second case here; only the case where $M_{N-1} < 0$ is of interest (if $M_{N-1} > 0$ then G_N is not dispatched and mark-up revenue is zero).

[12] The pdf of capacity margin is Normal on account of both available generation and load (in fact, a MOND) being Normal. Therefore M_{N-1} and M_N could conceivably be $-\infty$, though practically speaking the outer integral lower bound in (7.22) and (7.26) is set to $-1 \cdot [\text{maximum value of load}]$ (i.e., the highly unlikely situation when all generation is unavailable).

(5.17) and realised gross margins in the wholesale electricity market. In both cases these are given by (7.17) or (7.28), depending on market bidding assumptions.

To speed up the computation of (7.22)-(7.26) during the simulation, the outer integral is carried out using Gaussian quadrature (GQ), which requires fewer function evaluations than other methods, such as the recursive adaptive Simpson quadrature and is given by

$$\int_{-1}^{1} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

which after some manipulations, can be applied to the interval $[a, b]$. Here $n = 100$ is used.

7.3.5 Test of accuracy

To test the accuracy of this method, the results of a Monte Carlo (MC) simulation were compared with the MOND technique. That is, for the load (a MOND) and generator cost inputs given in Table 7.2, capacity mix given in Table 7.5 and mark-up function (5.8) calibrated to VOLL 10,000 £/MWh, random samples for load and generator availability were taken. Using these samples, the margin, M_N, price (5.11), and energy market gross margins were calculated for each unit. Close attention was paid to the tail of the distribution by using importance sampling. More precisely, high loads were oversampled using the component Normal of the MOND with the largest mean (i.e., row 1 in Table 7.2) with corresponding mean μ_1, standard deviation σ_1, and pdf $f^*(x)$. Samples were then weighted using the weighting function $W(x) = f(x)/f^*(x)$, where $f(x)$ is the 4 component MOND pdf and x is the random sample from $N(\mu_1, \sigma_1^2)$.

The MC simulation was repeated 10^6 times. The results of this test are displayed in Table 7.3. “MOND” shows the results for the MOND technique including the methods and approximations of Section 7.3, above. “Monte Carlo” is for a Monte Carlo (MC) test where available capacity is sampled from a Normal distribution with parameters defined by (2.16)-(2.17). Finally, “Bernoulli” samples available capacity from a Bernoulli distribution (i.e., 0 or full capacity based on uniform random number generation for given FORs). The Bernoulli test makes no approximations concerning the distribution of available capacity or its effects on prices and mark-up, and so is the standard against which the new MOND model should be compared. Comparing the MOND and Bernoulli tests shows the effect of using a Normal approximation for available capacity rather than the Bernoulli distribution. Encouragingly the MOND technique matched quite well to the MC simulation (with only a mild over-estimation of gross margins by the MOND technique). Further, the Normal approximation for available capacity also performs well. This test gives confidence that the MOND approximation is a good one. Note that the "energy expected per generator" in Table 7.3 for the Monte Carlo and Bernoulli tests is the result of multiplying the mean utilisation across the MC runs of each generator by the total theoretical available energy of each generator. For instance, if the capacity of generator $N-3$ is 11 GW and the mean utilisation is, say 0.75, then energy expected is given by 8760*11*0.75=72,270 GWh. For the MOND test, this is calculated using (7.11).

Loads (MOND)	μ_i (MW)	σ_i (MW)	p_i	
	43802	5743	0.43	
	41650	1410	0.12	
	33971	3372	0.33	
	26089	1771	0.12	
	$N-3$	$N-2$	$N-1$	N
Generator SRMCs (£/MWh)	7	40	45	60

Table 7.2: *Inputs for MOND test case.*

	Scarcity rent (£/MWh)				Price mark-up (£/MWh)				Energy expected per generator (TWh)			
	$N-3$	$N-2$	$N-1$	N	$N-3$	$N-2$	$N-1$	N	$N-3$	$N-2$	$N-1$	N
MOND	1.48	1.70	1.70	1.36	39.03	6.03	2.15	2.05	72.27	198.89	63.12	0.12
Monte Carlo	1.43	1.64	1.66	1.33	38.96	5.96	2.09	1.99	72.24	198.88	63.09	0.12
Bernoulli	1.41	1.61	1.63	1.30	38.91	5.92	2.04	1.94	72.24	199.40	63.11	0.12

Table 7.3: *Summary of results for Monte Carlo test.*

7.4 GB case study assumptions

The MOND technique is now applied within a simple dynamic investment model of the GB power system with the aim of providing insight into the investment risks associated with higher penetrations of wind power. Therefore the new dynamic model with 4 endogenous generator types (nuclear, coal, CCGT and OCGT) is applied to an 'energy-only' market setting with an initial capacity mix comparable to the GB power system and a VOLL of £10,000/MWh with a simulation time horizon of 30 years (2010-40).

The residual load facing thermal units for a particular hour is calculated in Section 6.5, i.e., by calculating hourly wind production using the base case wind capacity growth profile (Fig. 6.5) and subtracting it from the full empirical 2005-09 load data to create 30 residual load probability distributions (e.g., Fig. 6.6), one for each year. Each of the 30 year MOND residual load curves are precalculated assuming fixed underlying demand patterns and are then scaled over time in order to match the load growth assumptions. For instance, if average annual growth is expected to be 1.5%, then the MOND (7.8) after year t will be:

$$F_L(x) = \sum_{k=1}^{K} p_k \Phi_k(x|(1.015)^t \mu_k, \sigma_k). \tag{7.29}$$

Changes to the MOND standard deviation, σ_k, are not considered here.

Data on initial 2010 system capacity in Table 7.5 is derived by aggregating GB capacity data in Table 7.4 into the now four generator types and unit sizes described earlier. Once again, minor sources of peaking capacity such as oil and pumped storage are combined with OCGT and CHP and hydro are aggregated with CCGT plant to obtain the unit totals shown in Table 7.5. To reflect capacity already under construction in GB, 10.7 GW of CCGT capacity is assumed to come online during the first (1.5 GW), second (5 GW) and third (4.2 GW) years of the simulation. Note that the random variable used to represent unforeseen delays in the construction period during the real-time simulation (sub-section 5.4) is removed for this application. Existing plant included in the Large Combustion Plant Directive (LCPD) is modelled with a reduced lifetime based on the estimates of remaining generating hours given in [66]. All other existing units are given retirement dates consistent with the lifetime assumptions in Table 7.5. Table 7.6 shows the cost and financial input assumptions. The total annualised fixed costs per unforced MW and total interest accumulated during construction are shown in Table 7.7. To reflect restrictions on suitable sites for nuclear builds in GB, total installed nuclear capacity is constrained to 30 GW.

Is is assumed that there will be no net load growth during 2010-20. This is broadly in line with central Updated Energy Projections published by DECC [216] and base forecast winter peak demand figures from the GB SO [27]. Electricity demand after this point is assumed to grow at 1% per year.

Exchange rates are assumed to remain constant at €1.20/£ and $1.50/£. All calculations are carried out in real pounds sterling. Real discount rates are used owing to the forward estimates

Plant type	Capacity (GW)	# stations	# unit sets
Biomass	0.1	2	2
Coal	28.6	17	62
CCGT	27.5	39	118
CHP	1.8	9	19
Hydro	1.1	34	8[a]
Nuclear	10.9	9	22
Offshore wind	0.8	-	-
Oil	3.7	3	6
Onshore wind	2.2	-	-
OCGT	1.2	19	34
Pumped storage	2.7	4	16
Total	**80.6**	**136**	**287**

[a]Independent sets (i.e., not part of same hydro scheme).

Table 7.4: *Approximate GB installed transmission entry capacity in 2009. Appendix F of [27].*

Technology x **in (5.17)**	**Therm. eff.**	**FOR** ρ	**Lifetime** α **(yrs)**	**Build** τ **(yrs)**	**De-mothball** τ^D **(months)**	**Initial (GW)**	**No. of units**	**Unit size**
Nuclear	0.36	0.10[a]	40	7	-	11	22	500
Coal	0.35	0.14	40	5	6	27.5	55	500
CCGT	0.53	0.13	25	3	2	28.6	143	100
OCGT	0.39	0.10	40	2	2	7.7	154	50

[a]Recent years have shown a decline in the annual availability of the GB nuclear fleet (likely due to age), therefore this value is reduced to 75% for existing nuclear capacity. This estimate is the average of the aggregated energy output from installed GB nuclear capacity 02-09 [190]. New nuclear builds are expected 90% availability.

Table 7.5: *Generator technical assumptions and initial system capacity for GB case study with symbols defined in Section 5.3. Sources: [21, 26, 39].*

Technology x **in (5.17)**	**Capex £/kW**	**FC £/kW/yr**	**Var. O&M £/MWh**	**Equity return** κ	**Gearing** χ	**WACC** r^n **(nom.)**	**WACC** r^r **(real)**[a]
Nuclear	2,913	37.5	1.8	0.15	0.5	0.115	0.089
Coal	1,789	38.0	2.0	0.12	0.6	0.096	0.069
CCGT	718	15.0	2.2	0.12	0.6	0.096	0.069
OCGT	359	15.0	4.4	0.12	0.6	0.096	0.069

[a]Assuming a 2.5% rate of inflation.

Table 7.6: *Generator financial assumptions for GB case study with symbols defined in Section 5.3. Sources: [21, 26, 39].*

Technology x **in (5.17)**	**Annualised FC £/unfor.MW/yr**	**Total interest acc. during construction £/MW**
Nuclear	400,750	931,170
Coal	216,710	344,100
CCGT	91,840	96,030
OCGT	47,250	36,690

Table 7.7: *Generator (real) total fixed and interest cost assumptions for GB case study.*

for fuel and carbon prices being in real terms. All capital and operating costs are constant in real terms (2010 prices).

7.4.1 Intermediate analysis

The projected SRMC of the four generator types for a number of example years are shown in Table 7.8. Some example aggregate supply functions are shown in Fig 7.10 (without price mark-up). Note that coal and CCGT switch places in the system merit order in 2030 owing to the sharp increase in the ETS price after 2020. The fuel cost data behind these graphs is provided in Appendix A.4.

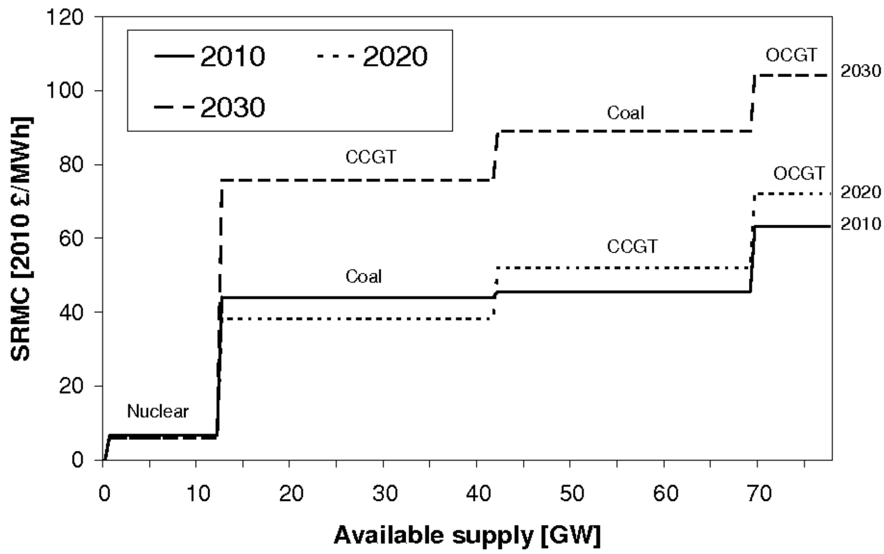


Figure 7.10: *Plot of stepped aggregate market supply function for example available conventional generation based on variable operating costs in Table 7.8.*

A plot of the technology screening curves is shown in Fig. 7.11. SRMC assumptions were for

£/MWh	Nuclear	Coal	CCGT	OCGT
2010	7	44	45	63
2015	7	37	49	68
2020	7	38	52	72
2025	7	64	64	88

Table 7.8: *SRMC assumptions for selected years. Note that uranium is modelled as constant at 70$/lb with a conversion of £1 = $1.5. Estimates in 2010 prices.*

2010 (Table 7.8) and annualised FC shown in Table 7.7. The plots show how coal is not optimal in the least cost solution in relation to other technologies, this situation did not improve for all levels of SRMC shown in Table 7.8. This is an important finding and will be highlighted again during the results interpretation.

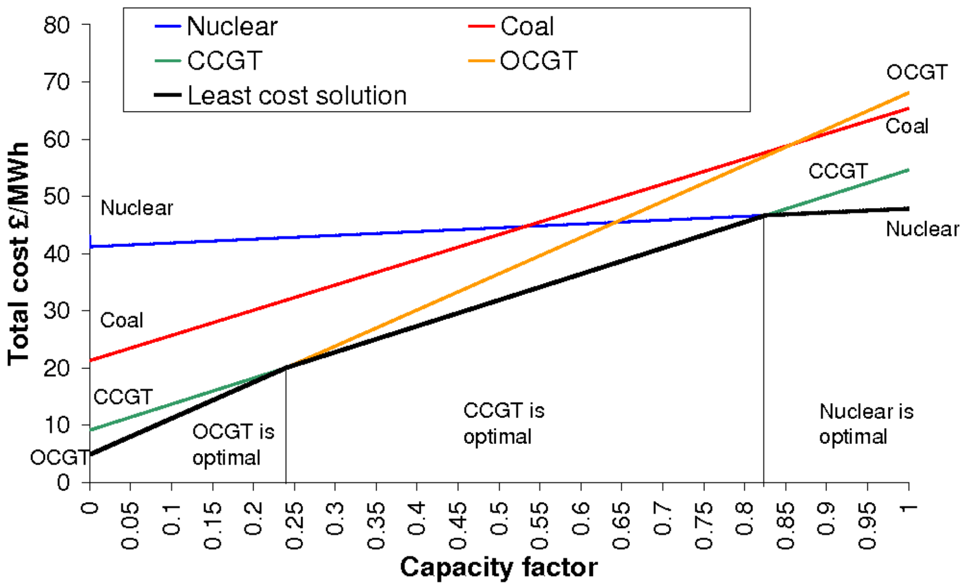


Figure 7.11: *Screening curves for technologies for 2010 cost data. Solid line indicates optimal (least cost) capacity mix.*

The effect of demand and wind capacity growth on load distribution percentiles over the period 2011-2025 is shown in Fig. 7.12. The 99.9%, 99.99% and 99.999% percentiles are shown, i.e., the loads which are exceeded approximately 9, 0.9 and 0.09 hours per year, respectively. One interesting thing to note about these plots is that although the wind penetration increases to over 30 GW, the reduction in residual peak demand percentiles is much less, which corresponds with the lower values of wind CC for high penetrations (cf. Fig. 6.10). Further, the MOND representation is less accurate for the tails of the distribution and so these demand levels should be interpreted with caution.

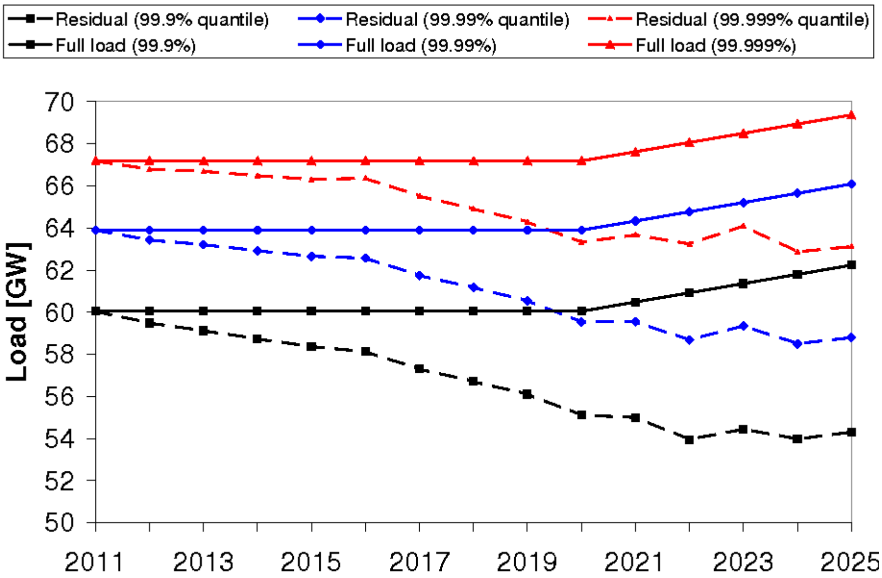


Figure 7.12: *Plot in change in demand percentiles for full and residual 2005-09 load data with demand growth described in section 7.2.*

7.5 GB Case study results

7.5.1 Base case results

For a Gaussian quadrature (7.29) with 100 points, the computational efficiency of the MOND technique allowed for each production costing run to execute in under 1.5 seconds. Thus the 7 stochastically simulated years and 100 MC simulations required 7x100x2=1400 seconds. Consequently, for the 30 year simulation, execution took between 525 and 1575 minutes, depending on the number of technologies chosen for investment. Plainly, reducing the number of GQ points shortens the production costing execution time (e.g., 25 points provided a 0.7 second saving per production costing run), though the accuracy of the MOND technique decreases. For instance, for the example shown in Table 7.3, the estimated price mark-up for generator $N-1$, calculated using (7.22) or (7.23), increases by 1.29 £/MWh. Likewise for generator N, calculated using (7.26), the estimate increases by 1.28 £/MWh.

Fig. 7.13 shows the evolution of total installed capacity in the simulation. New builds and plant retirements are shown in Fig. 7.14 along with the evolution of the mix and amount of generation over time in Fig. 7.15. Also shown in Fig. 7.13 is the simulated full and de-rated capacity margin (2.14). The FORs in Table 7.5 are used to de-rate conventional capacity, and for wind the long-term CC values plotted in Fig. 6.10. The forecast for peak demand is obtained from the 99.9% percentile of the year's MOND cdf for full load (Fig. 7.12).

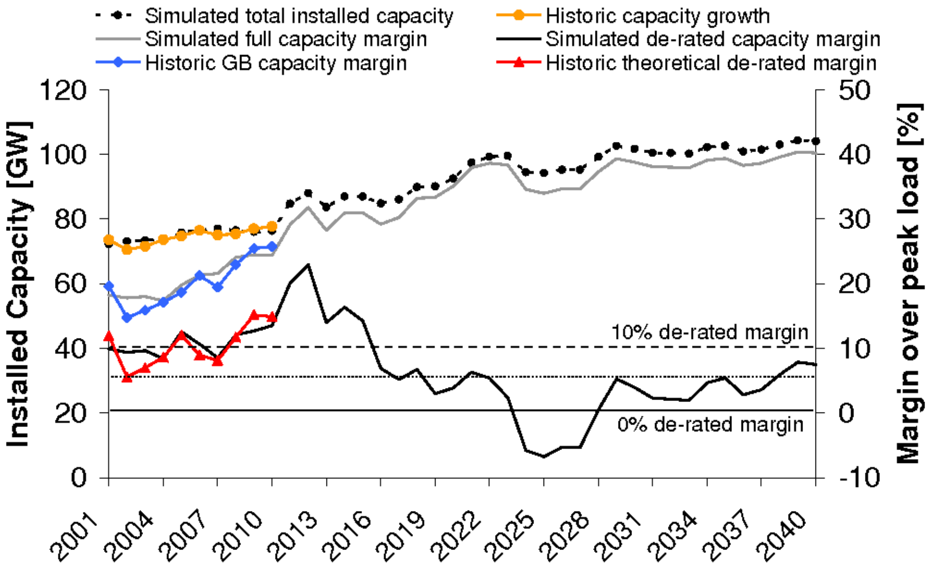


Figure 7.13: *Plot of simulated capacity growth, de-rated and full capacity margins. Theoretical historic GB capacity margin (2001-10) derived using data from [66] and [90].*

To compare the performance of the simulation against historic trends in GB, the historic simulation (2001-08) was re-run and a comparison between the modelled and actual capacity margins was performed (Tables 5.6 and 5.2 for cost and initial capacity data, respectively). The comparison (Fig. 7.13) shows that simulated margins do not perfectly match historic trends in all years (e.g., 2002/3 performs less well than in the initial implementation), yet there is a reasonably good agreement of the model with reality, which gives a degree of confidence in the realism of future projections. The average absolute difference between the historical theoretical de-rated margin and the simulation was 1.6% with a standard deviation of 1.3%.

The "historical theoretical de-rated margin" is as in Fig. 3.9. Furthermore, the comparison of available historical data on capacity additions plotted in Fig. 7.14 shows that the CCGT investments triggered by the simulation (columns after 2003) do not correspond in all years (e.g., 2008) but the volumes and timing are not unreasonably different. Further, 1.5 GW of coal-fired generation was chosen for investment in 2005, which is not true to reality (no new coal-

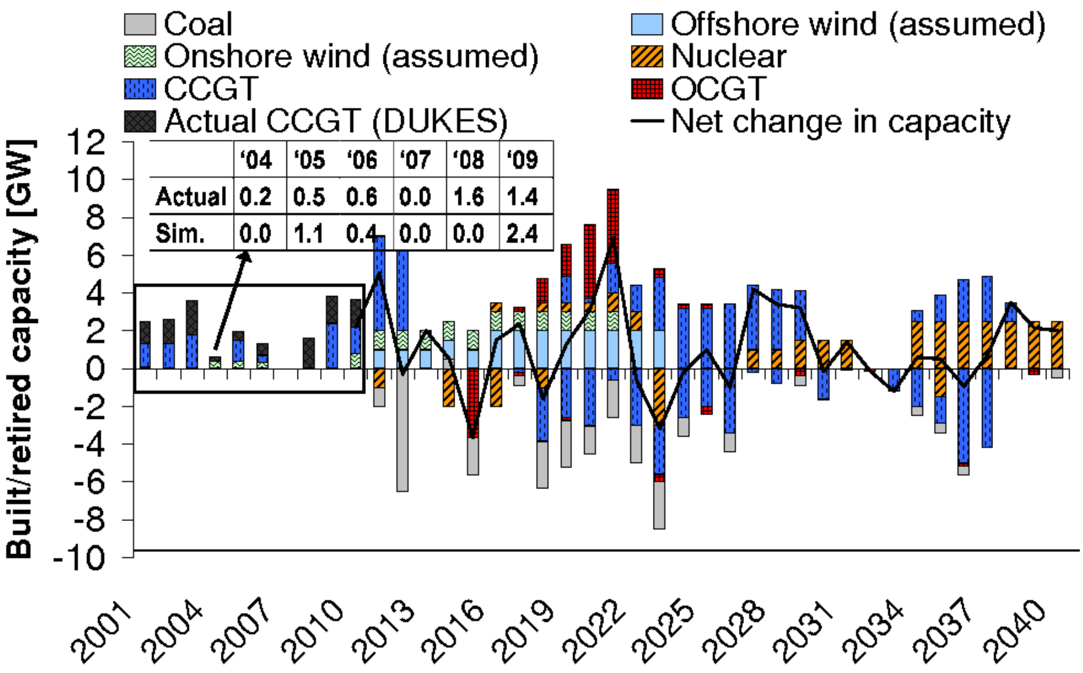


Figure 7.14: *Plot of simulated new builds and retirements over time. Negative bars indicate plant retirements and positive bars indicate new builds. Also shown are estimates for historic new builds (all CCGT) for 2001-09 (columns labelled 'CCGT DUKES' [190]).*

fired stations have been commissioned since 2000). However given that at least one generating firm in GB announced an intention to build a new coal-fired station (i.e., EOn at Kingsnorth), which was cancelled before construction began due to reasons beyond pure market rules (i.e., intense public and political lobbying), this result seems reasonable in a model such as this.

The future trend shows an erosion of de-rated capacity margins after around 2015. This coincides with the LCPD plant retirements and rapid offshore wind growth. Of the 30 simulated future years, the average de-rated margin is 5.6% with a standard deviation of 7.1%. De-rated margins are negative in 4 years, below 5% in 15 years, and below 10% in 25 years. For those years where margins are below 10%, an average shortfall of 1 GW of capacity was projected. The UK Government's recent consultation and subsequent white paper (July 2011) on GB electricity market reform has stipulated that a peak de-rated margin of 10% provides an acceptable level of generation adequacy risk [67]. Further, the GB SO has recently stipulated that a de-rated capacity margin of 5 GW over expected peak demand is desirable (see Appendix to [66]). By comparison these simulation results suggest that a lower than desirable level of adequacy

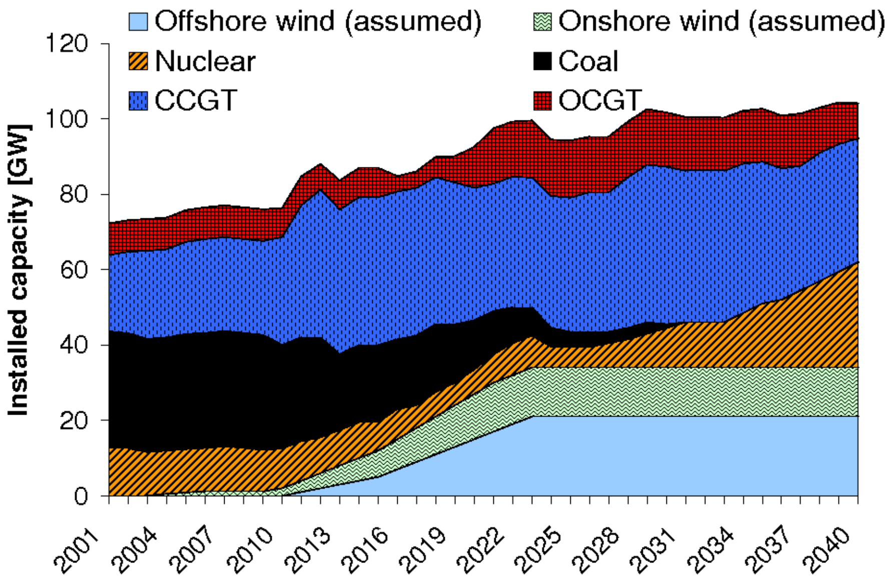


Figure 7.15: *Plot of total installed capacity over time, i.e., the result of the mix and amount of generation investment and retirements over time.*

risk could potentially occur.

The annual LOLE (7.15) and EEU (7.14) was also calculated. The average annual LOLE across the 30-year simulation was 0.03 hrs/yr (or 3 hours in 100 years) with a standard deviation of 0.05, and average annual EEU of 5.7 GWh (less than 0.002% of typical year's total annual energy demand). The yearly LOLE together with the volume of capacity required to meet a 5 GW de-rated capacity margin at peak is plotted in Fig. 7.16.[13] The graph shows how the LOLE is higher in some years, particularly in 2023-26, which is reflected in the de-rated peak margin shortfall. To put these figures in context, results from the historic generation adequacy risk calculation of Section 3.6 are recollected. The average 10 winter LOLE for 2001-10 was 0.06 hrs/yr with a standard deviation of 0.01 hrs/yr. This suggests that overall levels of risk do not appear to be significantly high relative to estimates of historic trends. However because the MOND is less accurate in the tails of the distribution the values for LOLE and EEU are used primarily to assess the changes in relative levels of risk over the simulation time horizon and to benchmark the sensitivity analyses described below rather than to predict absolute levels of system risk.

[13] A value of zero implies that de-rated margin is in excess of 5 GW.

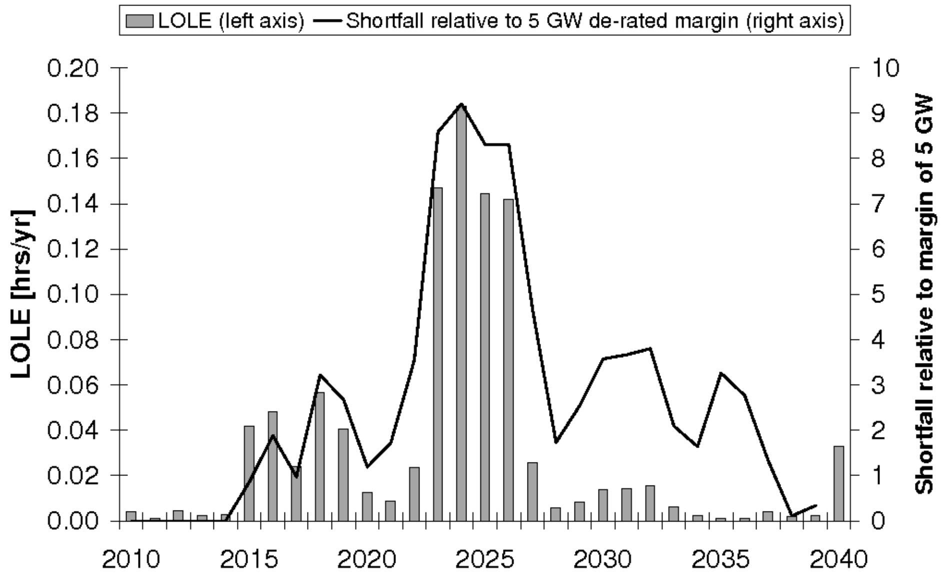


Figure 7.16: *Plot of simulated LOLE (bars) and capacity shortfall over 5 GW de-rated capacity margin (solid line).*

These projected risk and de-rated capacity margin figures suggest that the system may experience tight supply conditions at peak demand in some years. Some of these results can perhaps be explained by inspection of the residual load histograms from Fig. 6.6(d); the shape of the right-most tail suggests that even with very high penetrations, wind power does not contribute in all high demands periods. Furthermore the frequency of these high-demand/low-wind periods is too low to justify investment by private investors. And it is these very high-demand hours when the potential for a capacity shortfall is highest (excluding here SO actions such as voltage reductions). From a policy perspective, it is arguably uneconomical to design policies aimed at ensuring there is adequate generating resource available for these low-frequency events; an alternative approach could be to encourage demand-side participation through smart grids and smart metering. However these mechanisms will introduce price dynamics not currently witnessed in most liberalised energy markets and therefore careful consideration of the impact of demand response on generator's anticipated energy market revenues is required (cf. Section 9.1.5).

Further, an analysis of generator revenues shows symptoms of a boom and bust investment cycle. Simulated OCGT total gross margins, which include annualised capital costs (Table 7.7), are plotted in Fig. 7.17. Also shown are the triggered investments in this technology. Recall

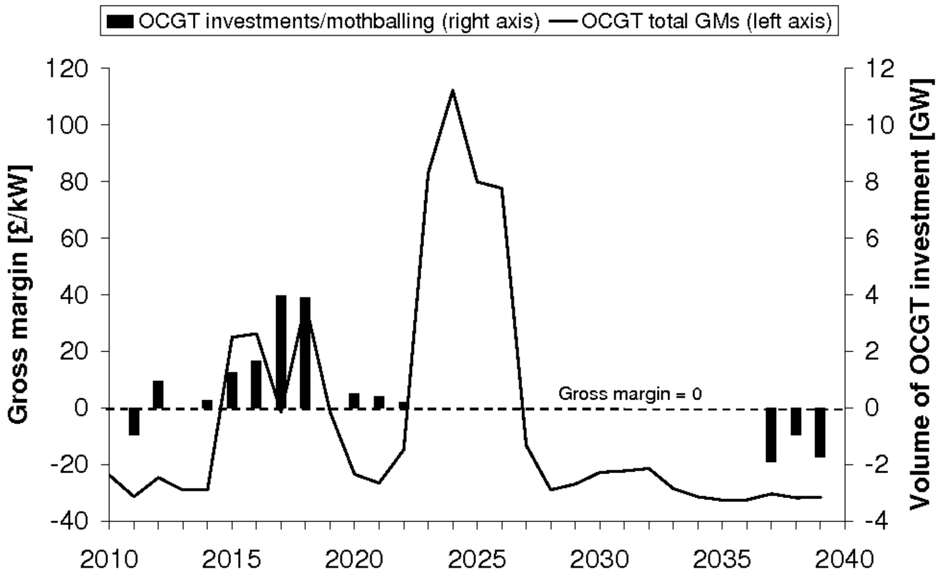


Figure 7.17: *Plot of simulated total gross margins for OCGT capacity (solid line, left axis). Also shown are the OCGT investment and mothballing amounts over time (black bars, right axis).*

here that investors stochastically simulate 7 years of prices, so the largest investment years (2017/18) include the forecast prices for (2023-25), the period when gross margins are highest. In contrast investment reduces in 2019-22 in response to expectations about prices being dampened (although not sufficiently to prevent an overshoot) as new investment in OCGT and other technologies enter the system. It is easy to see the pattern of high gross margins corresponding to those years where adequacy risk is highest (Fig. 7.16). The graph shows how the boom in OCGT investment in expectation of the high gross margins after 2023 is followed by a bust phase around 2026 when large volumes of new nuclear capacity begin entering the system (Fig. 7.15). This increases the capacity margin but reduces profitability for peaking units. In fact, a significant volume of plant is mothballed toward the end of the simulation time horizon suggesting that generators expect energy market revenues to remain low. It could be argued that this pattern of boom and bust investment would be less severe in a model that included an explicit representation of diverse agent behaviour. However as discussed in Section 7.2.2, the investment response of the model is smooth, as would be expected from a heterogeneous group of investors. Moreover, a model that considers firms with a portfolio of generation who might compute a joint revenue calculation across all their generation when considering a new investment may experience alternative investment dynamics than simulated here. Therefore, as with any simulation model that looks to the future, which is hugely uncertain, these results must be viewed with a degree of realism and caution. If the model were to be used in an actual policy analysis then extending the model to address these issues is a worthy topic for future research. This is discussed in Section 9.3 when recommendations for further work are presented.

The mothballing of OCGT early in the simulation (2011) is likely to be a direct result of expected high capacity margins out to 2014 (Fig. 7.13); the existing CCGT builds come online during these years in anticipation of LCPD-induced plant closures (Section 7.4). The profitability of other technologies is shown in Fig. 7.18. The plot shows similar profitability trends for CCGT compared with OCGT. Further, investment in OCGT capacity begins around 2014 with similar developments witnessed in CCGT and nuclear, in contrast no coal investments are made after 2005 as the plot of technology screening curves in Fig. 7.11 explains.

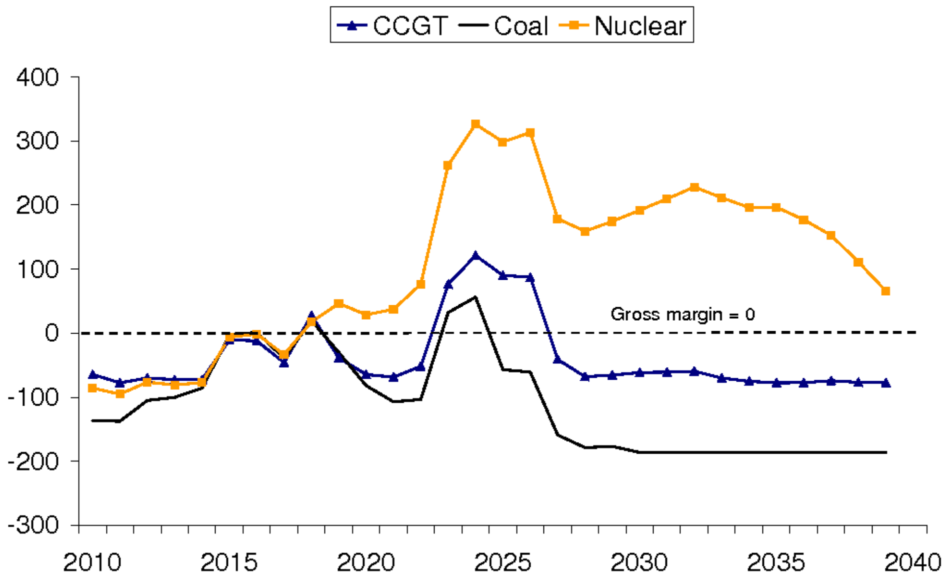
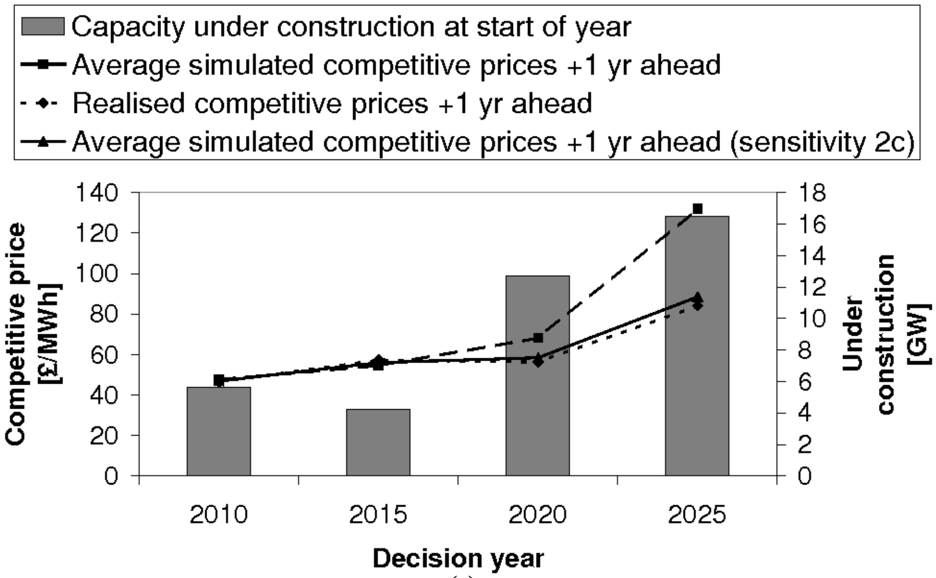


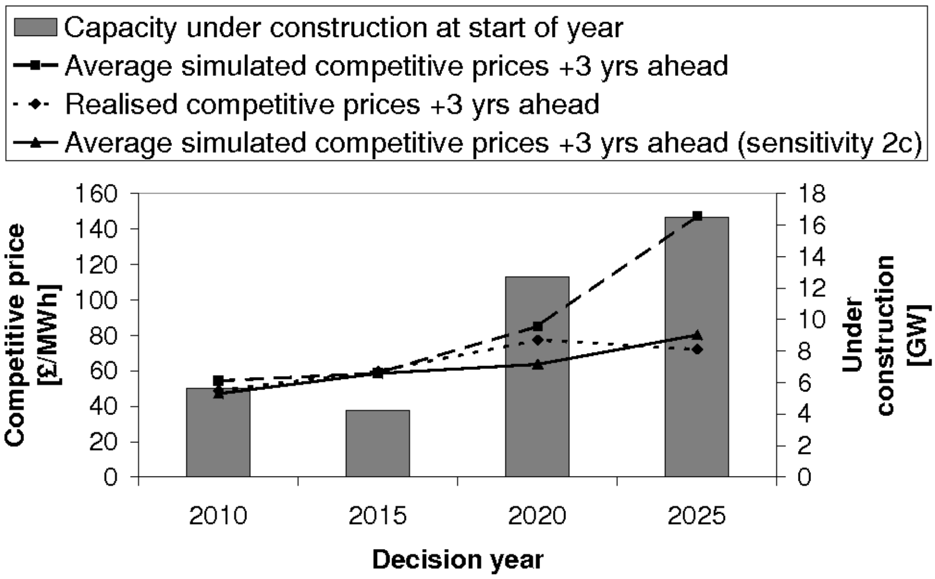
Figure 7.18: *Plot of simulated realised total gross margins for CCGT, coal and nuclear capacity.*

After 2025, very little CCGT or OCGT investment is triggered and endogenous capacity growth is minimal. Nuclear plants experience a sustained period of positive gross margins after 2023 (Fig 7.18), which is attributed to the rising SRMCs projects for fossil-fuel technologies (Fig 7.10) and hence increasing scarcity rents. This results in a period of intense and prolonged investment in nuclear after 2019 in response to high forecast gross margins. Combining this with the data presented in Fig. 7.14, implies a period of intense CCGT and OCGT investment for the 10 years after 2015 to offset retirements to existing capacity and respond to demand growth during 2020-25. Average annual endogenous capacity growth is -2.2% between 2015-25 as a result of 29.7 GW of new thermal build being offset by 42.2 GW of thermal plant retirements (options for lifetime extension were not considered here). This suggests that thermal capacity is not replaced on a like-for-like basis, which is hardly surprising given that average growth in installed wind generation is 12% over the same period. The ten years after 2025 provide better growth (average endogenous capacity growth 1.2% between 2026-35) on account of wind capacity levelling off, demand remaining flat and retirements continuing. In fact, nuclear is the only endogenous plant type to increase in terms of total installed capacity for the period 2025-40. This analysis suggests that new investments struggle to recover fixed costs during the period after 2026 owing to growth in nuclear capacity within a high wind system dampening energy market revenues for fossil-fuel generation.

An interesting analysis is to compare simulated real-time (annual) prices with investor expectations. This can be used to determine how well investors predictions of gross margins track those realised. Fig. 7.19 shows the average simulated competitive prices across the MC runs versus realised competitive prices for 1 and 3 years ahead for each of the years 2010, 2015, 2020 and 2025. Choosing 2020 as an example, in Fig. 7.19(a) (x-axis), the average of expected simulated competitive price for 2021 (1 year ahead) is higher than the realised price for 2021 (dashed line with diamonds) by 12 £/MWh. The degree of difference is also directly related to the volume of plant under construction; the more plant being built, the greater the over-estimation of market prices (and hence gross margins). Furthermore, the proportion of long lead time plant under construction exacerbates this trait. For instance, in 2020 the volumes of OCGT, CCGT and nuclear capacity under construction are 3.9, 5.8 and 3 GW, respectively, where as in 2025 it is 0, 10 and 6.5 GW, respectively, which is the year with the biggest difference. Note that, although mark-ups from market power are not shown here, precisely the same pattern occurs for simulated revenues from mark-ups. This can be traced back to the degree of uncertainty with which plant under construction is treated by the investor; the exact timings of when new plants will arrive are modelled as stochastic (sub-section 5.3.1). To test this hypothesis, a sensitivity case was modelled where investors have perfect foresight about investments in the pipeline, this is described in the next section (Test case 2c).



(a)



(b)

Figure 7.19: *(a) Plot of expected competitive market prices 1 year ahead from decision year (x-axis) and (b) expected competitive market prices 3 years ahead from decision year. Volume of capacity under construction at time decision is taken also shown in each case (columns).*

7.6 Sensitivity analyses

In order to test the robustness of the model, extensive sensitivity analyses have been performed on a number of model assumptions. Those found to be the most critical, some of which are plotted in Fig. 7.20 with key associated metrics shown in Tables 7.9 and 7.10, are described in the following sections.

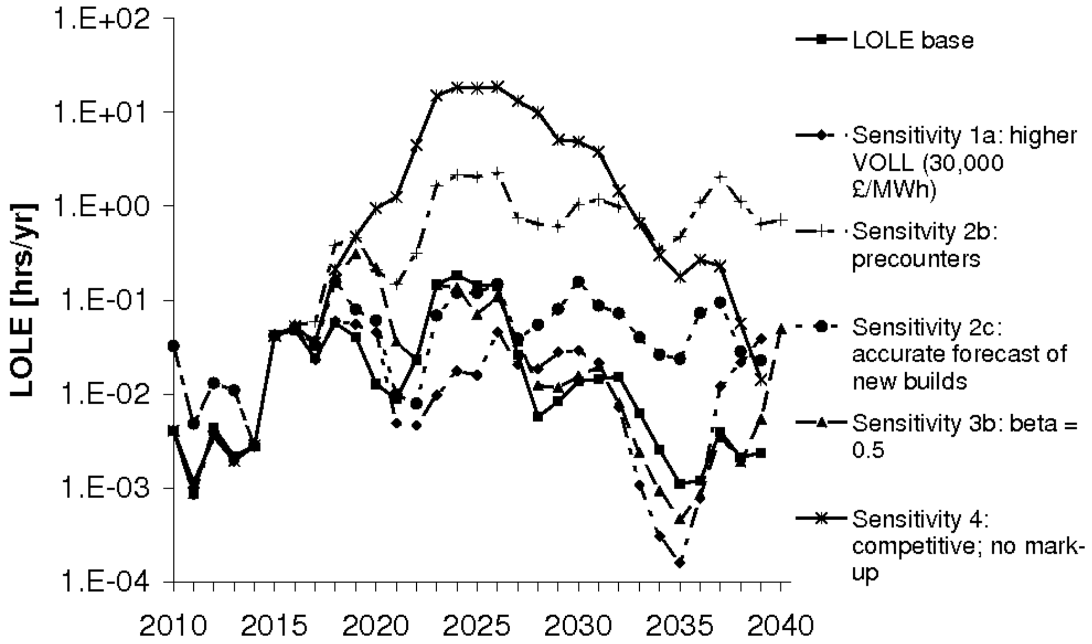


Figure 7.20: *Plot of simulated LOLE for selected sensitivities (test cases 1-6). Note the logarithmic scale. Also note once again that the MOND is less accurate in the tails of the distribution so the estimate for LOLE (7.15) and the forecast for peak demand (obtained from the 99.9% percentile of the year's MOND cdf for full load) will combine to likely under-estimate the true values of LOLE. Therefore these values are used primarily to assess the changes in relative levels of risk for the sensitivities listed.*

Overall, the model's qualitative behaviour was reasonable for the sensitivities listed and provided some useful insights, particularly when comparing the oligopolistic base case to the perfectly competitive market results (i.e., zero price mark-up, see below). Each of the experiments above is described in the following sub-sections.

7.6.1 Test case 1

The **level of expected scarcity price** (i.e., VOLL). Here, the maximum value of 10,000£/MWh is reduced and the price mark-up function $w(L, G_N^*)$ is altered (Fig. 7.8(b)). Experiments using case a) 'VOLL 30000': 30,000 £/MWh and b) 'VOLL 2000': 2,000 £/MWh are carried out.

	Security of supply risk metrics						Total investment and mothballing 2010-40 (GW)			
Test	Average de-rated margin 2010-40 (%)	Average diff. in margins relative to base case (%)	No. years negative	Deepest shortfall (%,yr)	LOLE (hrs/yr)	EEU (GWh)	Nuclear	Coal	CCGT	OCGT
Base case	5.6	-	4	(-6.8, 2024)	0.03	5.7	30.5	0	34	12.1
	(7.1)				(0.05)	(9.2)	-	(3.0)	(1.2)	(4.6)
VOLL 30000	6.5	1.2	1	(-1.3, 2025)	0.02	3.3	31.5	0	37	15
	(5.0)				(0.02)	(3.1)	-	(2.5)	(7.8)	(10.9)
VOLL 2000	3.8	-3.0	10	(-9.0, 2022)	0.22	46.8	30.5	0	32.6	13.3
	(7.8)				(0.49)	(116.7)	-	(3.0)	(3.0)	(5.0)
Believer	12.9	7.0	1	(-2.6, 2018)	0.01	1.9	30.5	0	34	32.4
	(6.7)				(0.02)	(4.8)	0	(3.5)	0	(15.0)
Pre counter	-5.1	-10.1	21	(-16.9, 2026)	0.71	158.7	30.5	0	27.6	12.0
	(8.5)				(0.70)	(167.3)	-	(3.0)	(3.6)	(5.3)
Accurate	1.7	-3.9	15	(-5.9, 2030)	0.06	10.2	31.0	0	24.2	22.9
	(4.5)				(0.05)	(8.3)	-	(1.5)	(3.6)	(6.2)
Resp. higher	6.3	0.8	6	(-4.7, 2024)	0.02	4.0	31.0	0	36.8	13.7
	(5.2)				(0.03)	(5.3)	-	(3.0)	(5.8)	(6.6)
Resp. lower	5.0	-0.5	8	(-5.8, 2023)	0.05	8.9	29.0	0.5	35.2	19.8
	(5.7)				(0.08)	(14.3)	-	(2.5)	(1.0)	(4.9)
No markup	-8.3	-13.0	20	(-33.2, 2024)	3.9	1474	31.0	0	32.6	8.5
	(14.4)				(6.25)	(2621.2)	-	(1.0)	(0.6)	(1.4)

Table 7.9: ***First summary table of security of supply and investment results for base case and sensitivity analysis. Figures in brackets below each case figure show standard deviation (security of supply risk metrics) or volume of plant mothballed (total investment and mothballing).***

Test	Security of supply risk metrics						Total investment and mothballing 2010-40 (GW)			
	Average de-rated margin 2010-40 (%)	Average diff. in margins relative to base case (%)	No. years negative	Deepest shortfall (%,yr)	LOLE (hrs/yr)	EEU (GWh)	Nuclear	Coal	CCGT	OCGT
Base case	5.6 (7.1)		4	(-6.8, 2024)	0.03 (0.05)	5.7 (9.2)	30.5 -	0 (3.0)	34 (1.2)	12.1 (4.6)
VaR higher	7.3 (5.2)	1.7	4	(-5.3, 2023)	0.02 (0.03)	3.4 (5.1)	29.5 -	0 (3.5)	30.6 (2.8)	18.2 (6.4)
VaR lower	6.2 (5.0)	0.9	5	(-5.7, 2023)	0.02 (0.03)	4.2 (6.1)	30 -	0 (3.0)	31 (3.6)	16.6 (7.1)
WACC lower	7.0 (4.8)	1.4	4	(-4.9, 2023)	0.02 (0.03)	3.3 (4.8)	30.5 -	0 (3.0)	28.4 (1.0)	7.8 (1.6)
WACC higher	4.8 (4.6)	-0.8	7	(-5.5, 2023)	0.03 (0.04)	5.8 (7.5)	31.5 -	0 (3.0)	39.2 (1.4)	19.1 (4.7)
Load SD	8.5 (5.3)	2.9	3	(-1.9, 2026)	0.01 (0.01)	2.0 (2.4)	30.5 0	0 (2.0)	32.4 (0.8)	22.6 (6.7)
AS reduce	3.9 (5.0)	-1.6	12	(-6.9, 2023)	0.05 (0.07)	9.3 (12.5)	31 -	0 (3.0)	39.2 (5.0)	9.9 (5.1)
FC increase	5.9 (4.3)	0.0	5	(-4.1, 2023)	0.04 (0.04)	7.4 (7.1)	30.5 -	0 (3.0)	38.8 (3.0)	12.8 (14.3)

Table 7.10: ***Second summary table of security of supply and investment results for base case and sensitivity analysis. Figures in brackets below each case figure show standard deviation (security of supply risk metrics) or volume of plant mothballed (total investment and mothballing).***

For test case 'VOLL 30000':, by increasing the VOLL, the expectation would be to see more investment as price spikes signals are stronger, which could potentially lead to a lower level of risk and also more investment relative to the base case. For test case 'VOLL 2000':, by reducing the VOLL, the expectation would be to see delays in investment as price spikes are dampened, which could potentially lead to a higher level of risk and also less investment relative to the base case. For instance, in a perfectly competitive market, in order for an OCGT plant to recover its fixed costs (47.25 £/unforced kW/yr), the price must reach the VOLL (10,000 £/MWh) in at least 4.7 hrs/yr (actually more to account for SRMC of production), and reducing the VOLL to 2,000 £/MWh increases this duration to 23.6 hrs/yr (Fig 7.21). The simulation results showed that increasing the VOLL to 30,000 £/MWh leads to a small improvement in average de-rated margins to 4.9% with a standard deviation of 5.0%. Furthermore, de-rated margins are negative during only 2026/27. The average annual LOLE reduces to 0.02 hrs/yr with average annaul EEU of 3.3 GWh. Interestingly, CCGT investment begins 1 year earlier and has a smoother cumulative profile relative to the base case (standard deviation of year-on-year volumes 0.8 GW versus 1.2 GW in base case). This displaces some OCGT investment early on, but overall volume during 2015-25 increase by 2 GW. Reducing the VOLL to 2,000 £/MWh alters investment timings, but overall volumes are only slightly reduced (3% lower than base case). The average annual LOLE increases to 0.22 hrs/yr with average annaul EEU of 46.8 GWh. CCGT investment starts 2 years later relative to the base case, which interestingly increases nuclear investment by 2 GW during 2015-20. OCGT investment is less intense during 2015-18 but volumes are significantly higher during 2020-23 (4.7 GW compared with 1.1 GW). This is in response to higher expected gross margins in 2022-26 as a result of less short lead time plant investment early on.

7.6.2 Test case 2

Investor expectations about new builds. Taking inspiration from [135] for, test case a) the *'believers'* case; here the investor ignores plants under construction when making expectations about revenues. The impact of new builds on prices are only considered once plants are fully operational (i.e., will systematically over-predict market prices). Secondly, case b), the *'pre counter'* is used where the investor views all plant under construction as operational (i.e., will systematically under-predict market prices). Finally, case c), a *'accurate'* investor is modelled

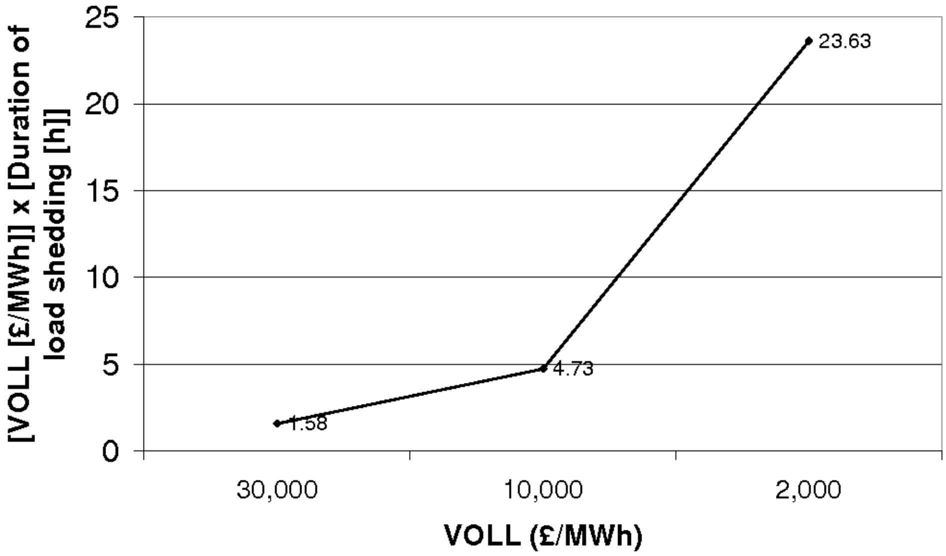


Figure 7.21: *Plot of required number of load shedding (or 'VOLL-priced') hours for OCGT generator to recover its fixed costs (assuming perfectly competitive market).*

where estimates about construction lead times match the delays experienced in reality (i.e., lead times shown in Table 7.5). This last case can be used to test the hypothesis that differences between realised and investor expectations about prices is due to the uncertainty surrounding capacity under construction.

For test case *'believers'*, the anticipation would be more investment relative to the base case due to the investor now systematically over-estimating revenues because they do not account for the impact of new builds when formulating price expectations. Results produced more investment than under the base case, with more severe over-shoot dynamics witnessed. As a result, average annual LOLE falls to 0.01 hrs/yr with generators unable to recover fixed costs with significant increases in the volume of mothballed OCGT capacity.

For test case *'pre counters'*, the anticipation would be less investment relative to the base case due to the investor now systematically under-estimating revenues because they include new builds before they are operational when formulating their price expectations. Results showed under-investment and average annual LOLE increases to 0.71 hrs/yr. As a result of less investment, overall profitability of existing plant improves, with some mothballing in the early part of the simulation in response to high forecast margins on account of plant already under construction at the start of the simulation (Section 7.4).

For test case *'accurate'*, the expectation would be reduced over-shoot dynamics and better prediction of realised prices relative to the base case. This is on account of the investor making a better prediction of the timing of capacity under construction, which is particularly important when there is a lot of long lead time plant (e.g., nuclear) under construction. Simulation results show that nuclear, CCGT and OCGT receive positive gross margins in almost all years after 2015, yet average LOLE doubles to 0.06 hrs/yr and with a noticeable oscillation of frequency 5 years and amplitude 0.12 hrs/yr (Fig. 7.20). Average EEU also increases to 10.2 GWh. Fig. 7.19 shows a significant improvement in investor price expectations at the 1 year ahead stage, with expectation 3 years ahead also more akin to reality. Expectations about years 4-7 ahead performed better than the base case, though due to the possibility of new investments (or mothballing) impacting on those years, significant differences remain.

7.6.3 Test case 3

The **aggregate investment response from the market**. This is achieved by altering the constant exponent, β, in the aggregate investment response function (7.4) to a) 0.9, 'Resp. higher' and b) 0.5,'Resp. lower' (Fig. 7.3).

For test case 'Resp. lower', the expectation would be that reductions to the exponent β in (7.4) to reduce investment, and increases under 'Resp. higher' to increase investment. This could also impact on the timing of investments later in the simulation on account of changes to investment levels early on. The results followed expectations; changes produced similar investment timings to the base case for OCGT and CCGT early in the time horizon but the reduced (resp. increased) levels of aggregate response lead to increases (resp. decreases) in generation adequacy risk in the medium-term (out to 2020) but higher (resp. lower) levels of investment later on in response to higher (resp. lower) forecast revenues. In the case of nuclear, a reduced response curve time-slipped the investment pattern by two years from 2020, whereas for the increased response curve the patterns were broadly similar.

7.6.4 Test case 4

The **ability of plant owners to exercise market power** and volumes of revenues received by doing so, reference 'No markup'. This is achieved by calculating expected gross margins using (7.17) instead of (7.28) (i.e., a perfectly competitive market versus the oligopolistic market in the base case).

For test case 'No markup', the perfectly competitive market case, the expectation would be lower overall levels of investment on account of price mark-ups being removed from the revenue calculation. Here more sustained periods of high prices (and hence scarcity rents) are required in order to recover invested capital and provide adequate return on investment. Indeed, when limiting the ability to exercise market power by removal of price mark-up, CCGT investments are most affected. Investment in this technology starts four years later than the base case and cumulative investment is on average 4.4 GW lower. Nuclear investments are also dampened, with investment during the first 15 years of the simulation 3 GW lower than the base case. These dynamics led to an order of magnitude increase in average annual LOLE and EEU to 3.9 hrs/yr and 1474 GWh, respectively. Of course these dynamics may differ in a model representing more sophisticated firms who account for the effect of their investment upon mark-ups for their existing fleet. In this case firms may deliberately not invest in order to keep prices (including mark-ups) high, however this type of investor logic is not considered here.

7.6.5 Test case 5

Investor risk preferences: achieved by case a) 'VaR higher': increasing critical q to 0.5 in VaR test (cf. sub-section 5.3.3), case b) 'VaR lower': lowering critical q to 0.01, case c) 'WACC lower': reducing investor WACC and case d) 'WACC higher': increasing investor WACC.

For test case 'VaR higher' and 'VaR lower', the expectation would be to see changes to the level of investments, particularly early on when the threshold for investment, which is considered only if the VaR criterion $p(V_x > 0) \geq (100 - q)\%$ is met. This means that an investment could be deemed attractive earlier (in the case where $q = 0.5$) or later (in the case where $q = 0.01$) relative to the base case. Moreover, the VaR criterion relates to the degree of risk aversion

(Section 5.3.3), therefore $q = 0.5$ models a less risk averse investor, and $q = 0.01$ models a more risk averse investor. Results of this experiment showed that modelling a less risk averse investor leads to marginally earlier nuclear investment (typically viewed as high risk due to high investment costs and operational inflexibility), with peaking plant investment timings remaining unchanged but volumes increasing in some years. Modelling a more risk averse investor leads to OCGT investment timings remaining unchanged but volumes increasing and displacing some CCGT (5.4 GW overall reduction in CCGT and 3.3 GW increase in OCGT); this is attributed to a higher degree of certainty about prices in the near term, resulting in less variance in projected revenues for short lead time plant. Hence the VaR criterion will tend to favour this type of plant for low values of q.

For test case ‘WACC lower’ and ‘WACC higher’, a similar expectation to that of ‘VaR higher’ and ‘VaR lower’ were held. Results show that the degree of impact that changes to the WACC had on investment timings was related to construction lead time. This is hardly surprising given that investment costs for long lead time plant are highly sensitive to the discount rate used. For instance, increasing the required rate of return on equity by 3% across all technologies delays nuclear investment by 3 years and overall volume by 5.5 GW during 2010-25. This leads to higher generation adequacy risk during 2018-22. Similarly, CCGT investments are delayed by 2 years although volumes are similar on account of the lower nuclear builds increasing expected gross margins and triggering investment. Reducing the required rate of return on equity by 3% leads to higher and earlier volumes of nuclear investment with annual average LOLE reduced, though gross margins are reduced.

7.6.6 Test case 6

Increasing investor uncertainty about load growth. Investors still consider load growth to be stochastic, however the standard deviation of the Normal distribution used to sample load growth is increased from 1% to 3%, reference ‘Load SD’.

For test case ‘Load SD’, the expectation would be that an increase to the standard deviation of load growth to lead to a higher standard deviation of the distribution of project value. This increased uncertainty surrounding profitability would reduced and/or delay investment. In fact, increasing investor load growth uncertainty leads to less CCGT and nuclear investment out to

2020, by comparison OCGT investment increases by 10 GW over the base case for the period 2019-21 in anticipation of high market prices in 2022-26. This high volume of investment lowers the generation adequacy risk and serves to dampen market prices, and OCGT gross margins are positive in 2026 only (the only year when de-rated margins are negative).

7.6.7 Test case 7

Economics of peaking plants: achieved by case a) 'AS reduce': reducing the amount of ancillary service (AS) revenue for OCGT plants and case b) 'FC increase': increasing total OCGT fixed costs (i.e., $TAFC_x$).

For test case 'AS reduce', a reduction in AS revenues (and hence profitability) was expected to reduce OCGT investments, although not massively on account of AS revenues alone not being sufficient to trigger investment (Section 7.2). In fact, reducing OCGT AS revenues from 10,000/MW/yr to 5,000/MW/yr alters investment timings and capacity choice early on in the simulation, yet long-term overall volumes are only marginally affected. More precisely, a reduction in OCGT investment during 2014-2018 (1.6 GW compared with 11 GW in base case) is somewhat offset by increases in CCGT (8.8 GW compared with 4.6 in base case) and nuclear (up 2.5 GW on base case) investment during this period. All three technologies were deemed profitable during this period; the model chooses the technology with the highest PI (5.21) and iterates until no additional plants are profitable (Section 7.2.2). A reduction in AS revenues for OCGT means that other technologies have more favourable profitability in the first iteration of the investment decision. OCGT is not chosen in subsequent iterations as a result of other capacity additions reducing its profitability to suboptimal levels. By choosing to invest in longer lead time plant, total LOLE over 2019-2023 increases from 0.29 to 0.97 hrs. Interestingly, higher volumes of OCGT investment occur in these later years (5.6 GW during 2019-2020 compared with 1.1 GW in base case). This is likely to be a consequence of the increased volumes of longer lead time plant under construction during this period increasing the differences between investor price predictions and reality (i.e., as discussed in sensitivity 2c).

Finally, for case 'FC increase', increasing OCGT $TAFC_x$ from 42.5 to 60 £/MW/yr (achieved by increasing capex from 359 to 510 £/kW) was expected to produce similar dynamics to case 'AS reduce'. This turned out to be the case, though the increase in LOLE is less severe (although still higher than the base case).

7.6.8 Test case 8

In addition to the VaR criterion, **Conditional Value at Risk (CVaR)** (cf. sub-section 4.4.2.3) with $q = 0.05$ is included as a sensitivity, reference 'CVaR'.

For test case 'CVaR', the expectation would be similar to test case 'VaR lower' on account of a more risk averse investor being modelled. Here investment is made if the expected project value is positive conditional on the revenue being below v_q, with q the confidence level, and investment is triggered if $E\left[V_x | V_x < v_q\right] > 0$, i.e., the worse $q\%$ of the 100 Monte Carlo simulated project NPVs determine the decision. The model produced some interesting results; nuclear investment is delayed by two years (first triggered in 2012), with lower volumes in the short-term (2.5 GW 2011-14, compared with 3.5 GW 2010-14 in base case) to medium-term (5 GW 2020-23 compared with 6.5 GW 2019-23), i.e., 30% overall reduction. After 2026 both timings and volumes were broadly in line with the base case. Further, OCGT investment was reduced by 18% and delayed (9.9 GW 2015-24 compared with 12.1 GW 2014-22 in base case), but CCGT investment experienced an increase of 14% (30.2 GW 2015-25 compared with 26.4 GW in base case). The average de-rated margin for the 30 simulated future years fell to 2.5%. The average annual LOLE across increased to 0.09 hrs/yr with a average annual EEU of 16.4 GWh. Again, 2023-28 was the period of highest risk.

7.6.9 Test case 9

To test the **importance of the primary fuel price projections**, a sensitivity where the DECC low coal (case a, reference 'low coal') and high (case b, reference 'high gas') and low (case c, reference 'low gas') gas price assumptions was tested. These are shown relative to the base case (central estimates) in Appendix A.4. Finally, a test (case d, reference 'carbon floor') with a 40 £/tCO_2 minimum (or "floor") ETS price was also tested.

For test case 'low coal', a reduction in the projected coal price (Fig. A.1(b)) would be expected to impact coal investment in the short-term only on account of the carbon price remaining high after 2020. Results showed this was indeed the case, yet the impact is less than expected with only 500 MW of coal triggered (in 2013), compared with zero in the base case. An unexpected result is that CCGT investment reduces to zero in 2015 and 2016 (1.4 GW and 0.2 GW, respectively, in base case), with OCGT increasing by 0.2 GW over the same period. This is largely unexplained, but is perhaps best attributed to the projected increase in the SRMC of coal occurring later in the simulation (Fig. A.4). This increase causes coal generation to move above CCGT in the merit order, thus increasing the projected scarcity rents (7.17) of CCGT generation. In contrast with this occurring later here than for the base case, the result is a reduction in CCGT investment early on in the simulation.

For test case 'high gas', a high gas price (Fig. A.1(a)) would be expected to delay CCGT and OCGT investments and perhaps trigger some coal and more nuclear investment due to increased forecast scarcity rents (7.17). Results show exactly this, although somewhat more so than expected. A sustained period of nuclear investment is experienced during 2010-19, with 15.5 GW of new investment triggered, compared with only 4.5 GW in the base case. Further, 8.5 GW of coal generation is triggered during 2012-18 with OCGT investment reduced significantly (5 GW 2013-23, compared with 12.1 GW in base case). Unsurprisingly, CCGT investment is also reduced, although some does still occur after 2017 (average 2.3 GW 2017-23, compared with 2.7 GW in base case), with overall volumes during the first 15 years of the simulation 42 % less. These findings correspond to the change witnessed for the least-cost of supply solution. More precisely, taking 2015 as an example, the technology screening curves are plotted in Fig 7.22. Notice how coal is now part of the least cost solution, which is not the case under any of the base case simulation years (e.g., Fig. 7.11). Interestingly, there is earlier and more intense mothballing of plants for both OCGT and CCGT; starting in 2027, 9 GW of CCGT and 6 GW of OCGT is mothballed. Some OCGT capacity is returned to service after 2034, although CCGT is not owing to most of this capacity reaching the end of its design lifetime before it is deemed profitable. The average de-rated margin for the 30 simulated future years rises to 6.7% with negative margins experienced in only 2 years. The average annual LOLE reduces to 0.02 hrs/yr with a average annual EEU of 2.9 GWh. The period of highest risk becomes 2018-20 (when negative margins occur), which is earlier than the base case. This is due to more longer lead time plant (i.e., coal and nuclear) being chosen for investment and hence the generation adequacy benefits of this capacity are experienced later in the simulation (e.g., no occurences of negative margins during 2024-27).

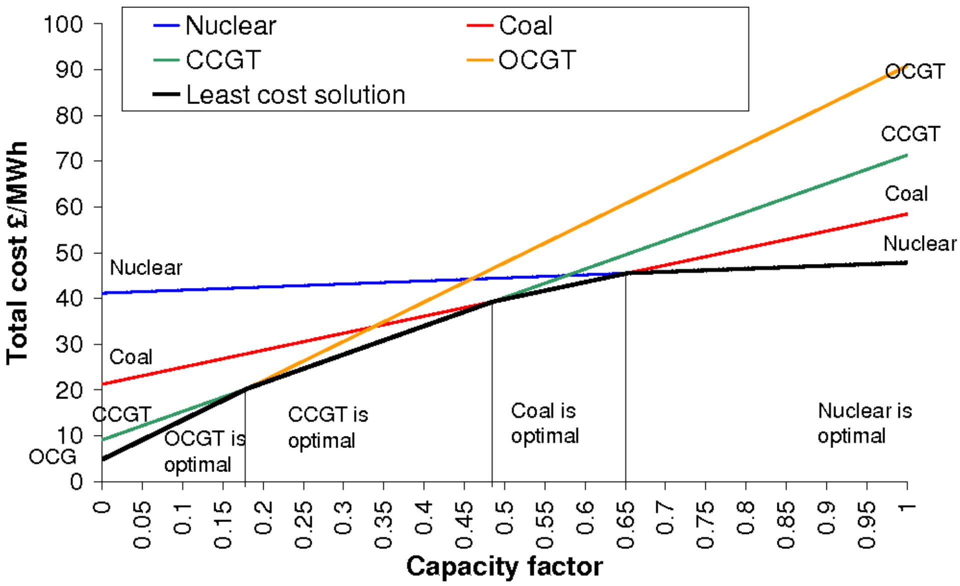


Figure 7.22: *Screening curves for technologies for 2015 SRMC cost data with high gas price sensitivity. Solid line indicates optimal (least cost) capacity mix.*

For test case ‘low gas’, the expectation would be that lower gas prices (Fig. A.1(a)) would lead to more gas-fired investment at the expense of nuclear generation. For a base-load generator, such as nuclear, in order to recover high capital and fixed costs, this technology requires large gross margins in the majority of hours. Therefore, reducing the SRMC of gas-fired generation (in the absence of coal investment), will reduce nuclear projected profitability. The results of the simulation show that no nuclear investment is triggered until 2027. Furthermore, 7.5 GW of coal-fired generation is mothballed during the first 5 years of the simulation, by comparison no OCGT mothballing occurs (Fig. 7.17). This leads to de-rated margins being on average 3% lower during this period. The average de-rated margin for the 30 simulated future years is 0.2% with negative margins experienced in a rather alarming 16/30 winters. The average annual LOLE increases to 0.07 hrs/yr with a average annual EEU of 12.9 GWh. The period of highest risk becomes 2015-26, which is earlier but lasts longer than the base case. Interestingly, CCGT investment remains largely unchanged, with only OCGT investments increasing (21 GW during 2014-23 compared with 12.1 GW in the base case). Fig. 7.23 shows how the capacity mix becomes dominated by gas-fired generation. By 2020, there is 36 GW and 14 GW of installed CCGT and OCGT capacity, respectively. This accounts for 53% of total installed capacity (75% of total thermal generation). Further, absolute levels of capacity are lower than in the base case in some years (average -2.8 GW less total installed capacity per year than base case during 2010-30), thus leading to more frequent and increased generation adequacy risk.

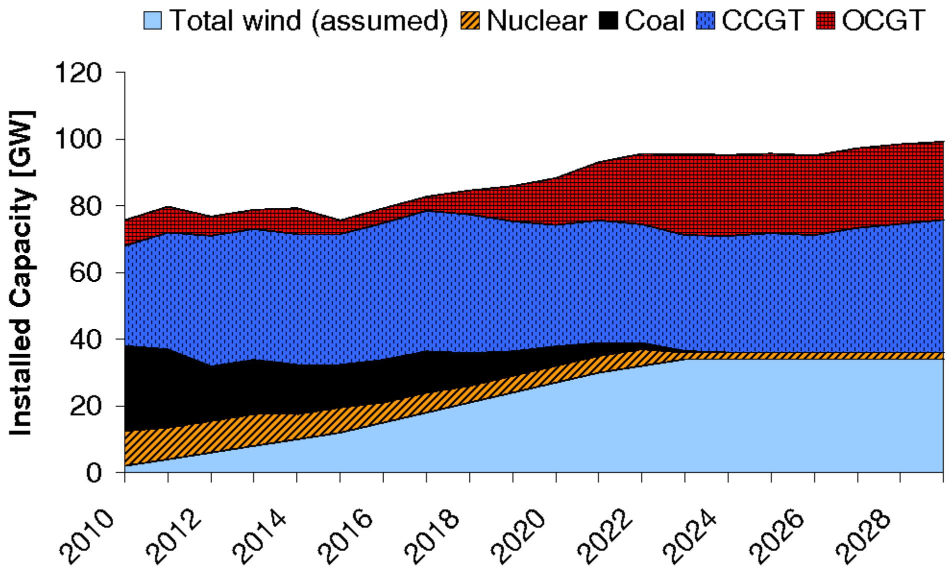


Figure 7.23: *Plot of total installed capacity for 2010-30 for test case 9c, i.e., the result of the mix and amount of generation investment and retirements over time.*

For test case 'carbon floor', the expectation about a 40 £/tCO_2 minimum ETS price would be that more nuclear investment occurs due to increasing gross margins (i.e., the opposite effect of case 'low gas'). This is indeed the case with levels of nuclear investment similar to that of test case 'high gas'; a sustained period of nuclear investment is experienced during 2010-21, with 13.5 GW of new investment triggered compared with only 4.5 GW in the base case. CCGT investment remains largely unchanged, in contrast OCGT investment reduces by 50%. Further, as with test case 'low gas', 4 GW of coal capacity is mothballed in the early part of the simulation. This capacity is not returned to service and is permanently retired in 2012 on account of the LCPD (cf. Table 5.4 for LCPD plant closure assumptions). A benefit of earlier and sustained nuclear investment is that the generation adequacy risk is reduced, although de-rated margins still remain low in 2023-26. The average annual LOLE reduces to 0.01 hrs/yr with a average annual EEU of 2.3 GWh.

7.6.10 Summary of test results

In summary, a pattern of increased levels of risk relative to the historic calculation and erosion of de-rated capacity margins was experienced during the 2020s to some degree in all cases. Furthermore, the period of highest security of supply risk is 2023-28, though the magnitude of the risk (measured by LOLE) differs between experiments, with the high gas price sensitivity bringing the riskiest period forward to 2018-20. A prolonged period of increased security of supply risk is experienced throughout the 2020s for the perfectly competitive market case. Also, providing investors with perfect foresight about capacity under construction produces less investment; this leads to more frequent periods of relatively high LOLE and low de-rated margins after 2020. However generators experience positive gross margins in more years on account of reduced surplus margins, and hence higher prices.

The 2020s is the period of most intense change; during this decade over 40 GW of new capacity is built with over 34 GW retiring; 14.5 GW of capacity leaves the system during the period 2023-26 alone. This further exacerbates the higher relative levels of risk during this period. Furthermore, no new coal investments are made, with the majority of new investment coming from gas-fired generation. A sensitivity on the primary fuel prices showed that lower coal prices (or indeed higher gas prices) may lead to different technology choice, with relative levels of high generation adequacy risk shifting to earlier in the time horizon.

7.7 Chapter summary

This chapter provided details of the updated dynamic investment model. New features of the investment decision element of the model are provided in Section 7.2, including how investors account for an increasing wind penetration and the introduction of an aggregate investor response curve. In Section 7.3 the MOND was formally defined and its application to a market situation is given. This included a novel derivation to account for revenues from price mark-ups and a test of MOND technique accuracy. In Section 7.4 described input assumptions for the GB market simulation and the wind models used from Chapter 6. Results from the dynamic simulation of capacity investment in a system with an existing installed capacity similar to that of GB is given in Section 7.5, and an extensive sensitivity analysis followed in Section 7.6. Results showed a pattern of increased levels of risk relative to the historic calculation and erosion of de-rated capacity margins was experienced during the 2020s to some degree in all cases. The topic of the next chapter will be to determine whether explicit capacity mechanisms can be designed to alleviate resource shortfall and prevent investment overshoot. The results here indicate that such a mechanism may be desirable to improve reserve margins in the mid 2020s.

Chapter 8
Implementation of a Capacity Mechanism

In this chapter, the scope for including a generation capacity investment control signal to dampen investment cycles is explored. This begins with a discussion on the challenges faced when applying techniques from classical control theory in order to design a robust investment market controller for the dynamic model, and how reverting to existing modelling methodologies can still provide new insights. These insights are underpinned by a discussion on the need and design for a capacity mechanism for the GB market test case based on the results of Chapter 7. Two capacity mechanisms are tested, these are a strategic reserve tender and a market-wide capacity market, and their impact on relative levels of risk is analysed.

8.1 Introduction

In dynamic control, a *robust controller* is defined as a control mechanism that minimises the effect of uncertainties whilst achieving the desired level of control over the system. For the control system depicted in Fig. 5.1, the aim of the controller is to dampen investment cycles and maintain the suitable level of security of supply risk. Initially, it was envisaged that the measurement signal would be full or de-rated capacity margin, and the controller would behave like a capacity mechanism by increasing (or reducing) additional generator revenue streams if capacity margins (a proxy for security of supply risk) fall below (or rise above) the reference point, e.g., Fig. 8.1. In a system with a high wind penetration, this reference point would be a target level of de-rated capacity margin, say 10%.

In order to design a robust controller, all elements of the system must be represented using differential equations. This is termed the *state-space* representation. An example of a state-space representation for installed capacity is shown in (5.1). Here the state variable, x, is the installed capacity at month m, and y is the system output, which is the sum of exogenous (uncertainty in supply and project abandonment) and endogenous (new build or mothballed

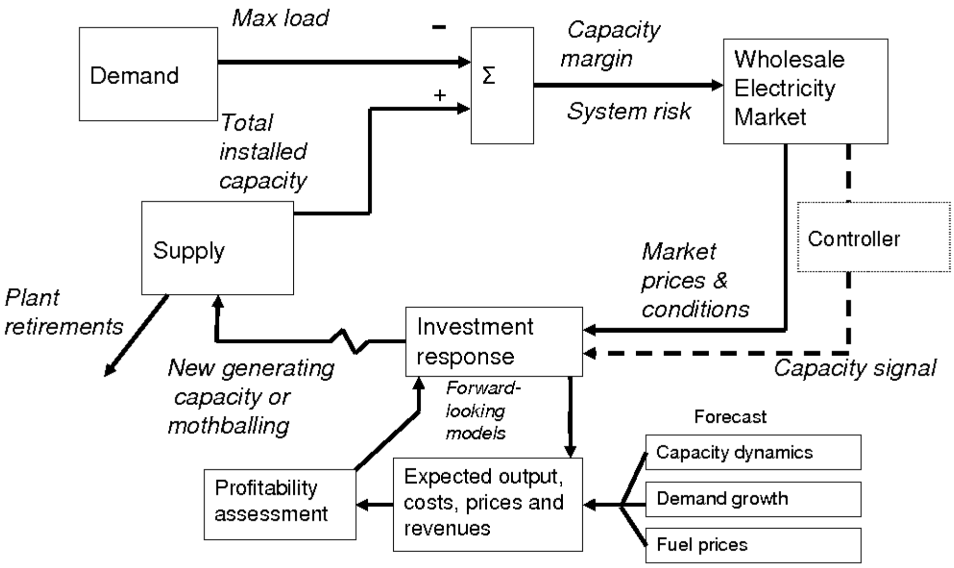


Figure 8.1: *Electricity investment as a control problem with the proposed controller; the aim of the controller is maintain a reference level capacity margin by making generator revenues more predictable.*

capacity) variables. In this case, a challenge arises for the system depicted Fig. 8.1 due to the investment block being a "black box", i.e., it cannot be represented using differential equations. In plain terms, the aggregate response of investment to market conditions can be measured, yet unlike the dynamics of capacity change (e.g., (5.3)), the exact method by which the decision is carried out cannot be represented using differential equations.

In the absence of a state-space representation, *system identification* can be used to created a mathematical model of a dynamic system with "black box" characteristics using observed input-output data. For instance, the Matlab System Identification Toolbox [225] can be used to fit linear and nonlinear models to observed data (e.g., in this application the simulated investment response or de-rated capacity margin). These model structures include state-space models which could then be used to design a robust controller. However this functionality was only surveyed during the final stages of research and, given the limited time available in which to absorb and then apply these tools, it was decided to incorporate a capacity mechanism into the dynamic model using more established methods of implementation. That said, the exact specifications of the capacity mechanisms implemented here are unique. The approach taken to including these elements in the dynamic model are outlined and tested in the following sections.

8.2 A GB market capacity mechanism

Broadly speaking, any debate about whether or not a capacity mechanism is required must be informed by a robust generation adequacy risk measure. This permits assessment of the risk in any given investment scenario, and also guides the design of energy and capacity markets by setting a baseline for what the market should deliver (such as an adequate capacity mix consistent with appropriate returns on investment). In the following sections, drawing on previous model results and underlying input assumptions for generation reliability, the need and design for a capacity market in GB is explored.

8.2.1 Need

The results presented in Sections 7.5 and 7.6 suggest that the GB system may experience increased generation adequacy risk during the mid to late 2020s. This motivates the consideration of capacity mechanisms, which can be employed to make generator revenue streams more predictable and thus bring on more timely investment. Fig. 8.2 shows the social optimal level of unserved load in terms of cost and reliability using the simple model of reliability (sub-section 2.4.3). The optimal level (circles) for each VOLL can be determined using (2.8). In the main simulation, the average 30-year EEU in the base case was 5.7 GWh per year, suggesting the system may fall below the optimal level of reliability for a VOLL of 10,000 and 30,000 £/MWh. Furthermore, during the period of highest risk (2023-28), the maximum EEU was 33.4 GWh per year, suggesting that this period is less reliable than even the optimal duration for a VOLL of 2,000 £/MWh. That said, these figures of EEU are calculated using the MOND technique so cannot be considered altogether reliable in terms of absolute values. Therefore, when designing a capacity mechanism, the EEU should be used to compare relative change between experiments (Section 7.6), but owing to the possible inaccuracies of the estimate, cannot be relied upon to produce a robust generation adequacy risk measure. With this in mind, attention turns to a proxy for system risk, namely the de-rated capacity margin.

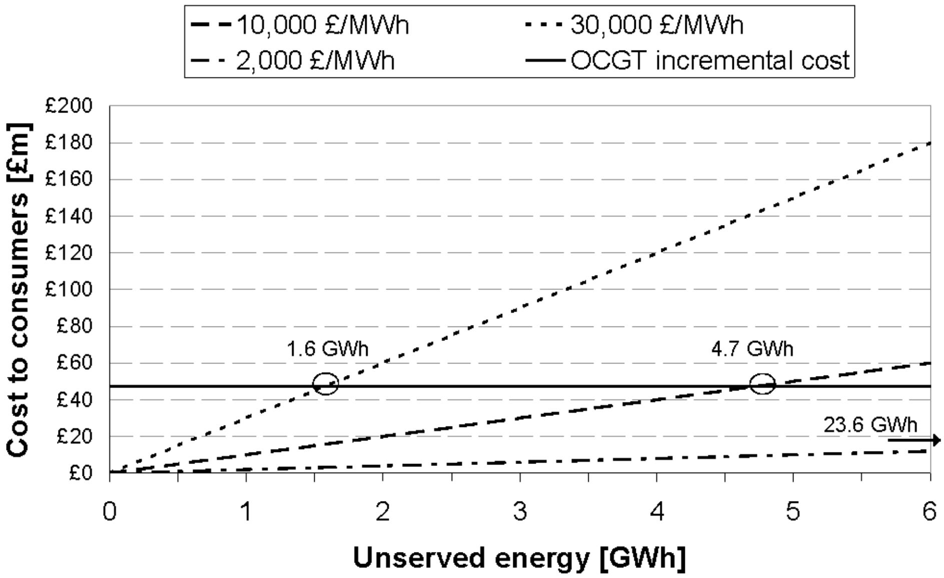


Figure 8.2: *Plot of optimal level of unserved energy for different levels of VOLL. Based on incremental cost of an OCGT plant with FC of 47.25 £/unforced kW/yr (cf. Table 7.5). Note the £million scale applied to the y-axis.*

8.2.1.1 De-rated capacity margin

The control problem formulation allows for the de-rated capacity margin (DRM) (2.14) to be observed, yet this signal is dependent on the scaling factors (i.e., FORs for conventional generation, and CC for wind) applied to each generator type. Owing to the lack of available data on plant technical availability in GB, the FORs applied to de-rate capacity use best available estimates (e.g., [26, 65]) and must be viewed with a degree of caution. Furthermore, following the discussion on wind CC results in Chapter 6, the same can be said of the availability of the wind resource at periods of highest demand. Therefore, before proceeding to the design of a capacity mechanism, it is worthwhile considering the sensitivity of investment scenarios to the underlying input assumptions for expected generator reliability.

Fig. 8.3 shows the level of realised DRM for the base case investment scenario when the FORs in Table 7.5 are reduced (values shown on legend). Also shown is a low-wind scenario where the capacity credits for wind shown in Fig. 6.10 are calibrated in order to provide a capacity credit of 5% for the high penetrations (Fig. 8.4). Plainly the DRM is on average 1.3% below the base case measurement when nuclear FOR is increased from 0.1 to 0.2, 5.9% below when CCGT FOR is increased from 0.13 to 0.2, and 2.7% below when the CC of wind are reduced.

For conventional generation, the reduction is dependent on the level of installed capacity for each tecnology. For instance, the reduction in de-rated margin is higher when increasing the FOR of nuclear later on in the simulation because installed capacity is near to 30 GW, however it has little impact during the period of lowest DRMs when installed nuclear capacity is low (average of 6.6 GW over 2023-28).

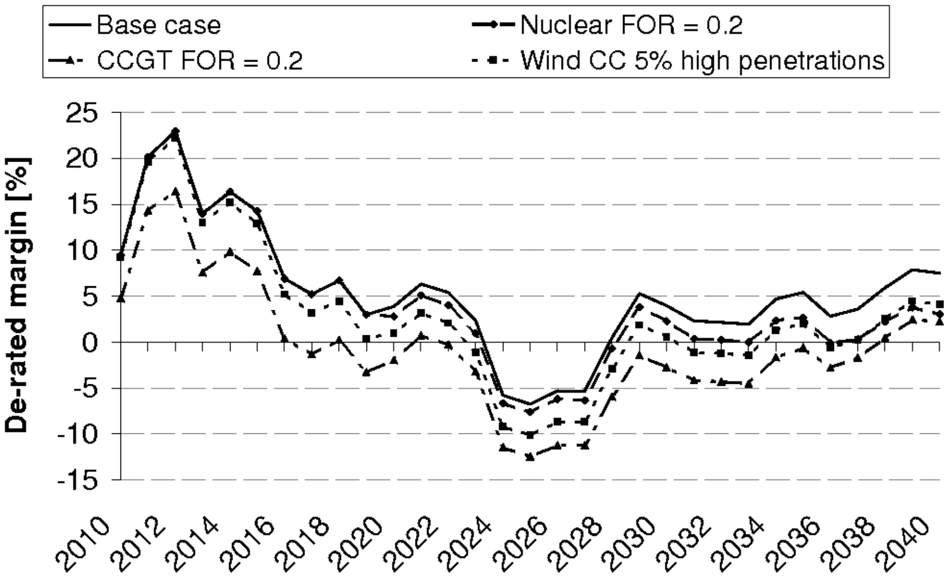


Figure 8.3: *Sensitivity of measured de-rated margin to assumptions about de-rating factors.*

This simple analysis demonstrates the impact that the assumed availability of supply will have on measured de-rated margins, and although not explicitly included here, a similar pattern occurs when uncertainty surrounding peak demand is tested (e.g., sampling quantiles in excess of 99.9%; Fig 7.12). The importance of this analysis is further evident when one considers predicting the de-rated capacity margin in advance of realised levels of supply and demand, which is particularly important for many of the capacity mechanisms that exist presently (e.g, PJM's capacity market).

8.2.2 Design and testing

Drawing on the observation in Section 7.5, that according to the UK Government, a peak de-rated margin of 10% provides an acceptable level of generation adequacy risk [67], and furthermore, that views on the use of either a strategic reserve tender or market-wide mechanism in

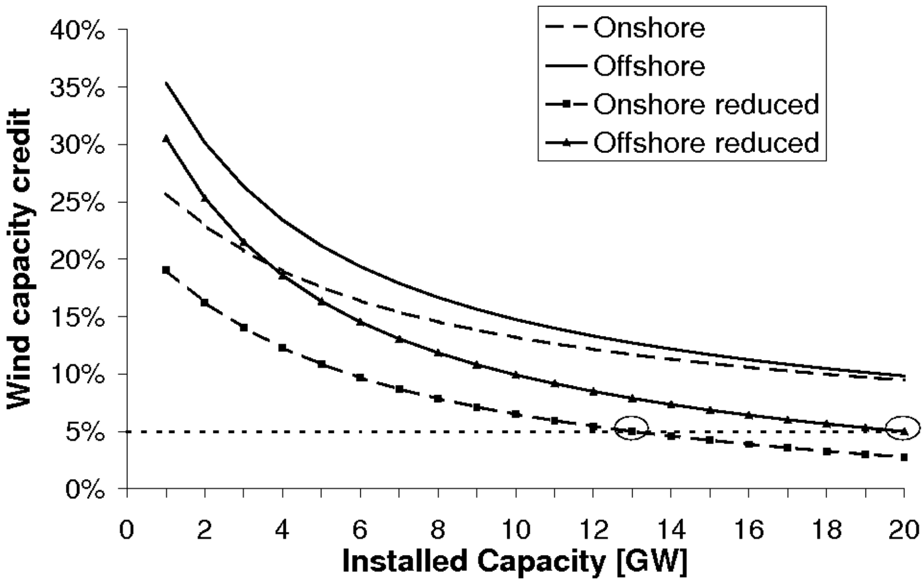


Figure 8.4: *Altered wind capacity credit graph. Calibrated so CC is 5% for the highest capacity penetration; calibration point highlighted with circles (onshore 13 GW, offshore 20 GW).*

the form of a capacity market [226] are sought, two capacity mechanisms are tested within the model. Each has the aim of achieving a 10% de-rated capacity margin. They are a *strategic reserve tender* and a *market-wide capacity market*. Note that both the mechanisms tested here are quite simplistic compared with existing capacity markets, yet these implementations can provide insights into the effectiveness of capactiy mechanisms in mitigating generation adequacy risk concerns in a system with a high wind penetration, an area of research which is much less understood.

Before describing the details of each mechanism, the time horizon over which the capacity mechanisms operate must be defined. Some capacity mechanisms involve real-time operation (e.g., E&W Pool capacity payment) and others annual (e.g., PJM's capacity market operates from June 1 to May 31 [227]). Each of the mechanisms tested here are forward-looking, that is a forecast of the de-rated margin a number of years into the future is made and capacity is procured, or a capacity price is determined, based on this projection.

Using the same notation as (5.3), the forecast for total installed capacity of plant type x for year

$t+T$, made at year t, can be obtained by

$$I_x(t+T) = I_x(0) + \sum_{i \in A_x} \xi_j^B - \sum_{i \in R_x} \xi_j^B + \sum_{i=1}^{n_x^M} \xi_i^M \tag{8.1}$$

where A_x is the set of all plant expected to be built by year $y+T$: $A_x = \left\{ j | t+T \geq \eta_j^B + \tau_x \right\}$, and similarly R_x is the set of all expected retired plant: $R_x = \left\{ j | t+T \geq \eta_j^B + \tau_x + \alpha_x \right\}$. Thus the forecast expected de-rated margin (FDRM) for year $t+T$ made at year t is given by

$$FCDM(t,T) = \frac{\sum_{x=1}^{4} I_x(t+T) + DRW(t+T)}{MPPD(t+T)} - 1 \tag{8.2}$$

where $MPPD(t+T)$ is the forecast for most probable peak demand and $DRW(t+T)$ is the forecast de-rated wind capacity (ELCC); both are exogenous parameters to the dynamic investment model. $DRW(t+T)$ is obtained from a combination of Fig. 6.5 (installed on and offshore capacities) and Fig. 6.10 (capacity credits used to de-rate wind). The forecast for peak demand, $MPPD(t+T)$, is obtained from the 99.9% percentile of year $t+T$'s MOND cdf for full load (Fig. 7.12). Thus, the previous caveat on intepretation of absolute values applies to the FCDM as well.

When planning a system, lead times due to engineering, procurement and construction must be accounted for. For instance, in the GB case study presented in Section 7.4, if a central estimate for de-rated capacity margins taken at year t, for delivery in year $t+3$ was required, only new investments in OCGT plant currently under construction would impact on the level of generation adequacy actually experienced in year $t+3$. By comparison if looking ahead to year $t+8$, all new investments in year t will need to be considered.

8.2.3 Strategic reserve tender

The analysis begins with a strategic reserve tender (SRT). Taking the definition in [226], this mechanism consists of a centrally-procured volume of capacity *"which is removed from the electricity market and only utilised in certain circumstances"*. These circumstances are likely to occur when either all other available resources have been exhausted or the wholesale price reaches a politically unacceptable level (i.e., above the perceived long-run marginal cost of the most expensive plant type, but below the VOLL).

To allow for construction lead time, capacity is procured three years ahead of required delivery and is restricted to OCGT plant only. Three years is chosen on the basis that this will leave sufficient time for any required OCGT capacity to be built and, if investments in other technologies are made in the meantime, they will not influence the level of generation adequacy experienced three years out, but will contribute in later years.

Capacity is procured on the basis of meeting the 10% de-rated target reserve margin. If no shortfall is expected, then no capacity is procured. To account for the fact that strategic reserve, like all other OCGT capacity, is subject to forced outages, ρ, the required tender volume (RTV) is determined by scaling the forecast required de-rated capacity in year $t + T$

$$RDC(t + T) = max\left[0, MPPD(t + T) \cdot (0.1 - FCDM(t, T))\right], \tag{8.3}$$

by the expected FOR of an OCGT plant, i.e., $RTV = RDC/(1 - \rho_{OCGT})$.

The modelled investor is aware of the reserve tender, but is unsure exactly what price will activate the reserve capacity. Further, the investor assumes that strategic reserve capacity will have the same impact on wholesale prices as any other volume of capacity entering the system (i.e., it will dampen prices) with the additional SRT resource combined with other OCGT peaking resources when calculating projected energy market revenues (7.28). This logic is extended to the decision to mothball plant (7.3), though it is restricted to capacity which is built as a result of energy market and AS gross margins alone, i.e., not part of the SRT. All SRT is assumed to be paid an availability and utilisation price adequate to provide a reasonable rate of return on investment. Finally, in each decision year, the investment decision is run first and the required SRT capacity calculated afterwards, thus giving the market every opportunity to provide the required capacity in order to meet the 10% target de-rated margin.

8.2.3.1 Simulation results

Fig. 8.5 shows the evolution of de-rated capacity margin over the period 2016-40 and that the period of highest adequacy risk in the base case was 2023-28. The graph shows how de-rated margins improve significantly relative to the base case. Of the 30 simulated future years, the average de-rated margin is 12.9% with a standard deviation of 4.5%, relative to the base case where equivalent figures of 5.6% and 7.1%, respectively, were observed. The 10% target DRM is achieved in 21 of 30 simulated years (compared with 5 of 30 years in the base case). The average annual LOLE across the 30-year simulation was 0.002 hrs/yr with a standard deviation of 0.003, and average annual EEU of 0.353 GWh per year. These figures are an order of magnitude lower relative to the base case projections.

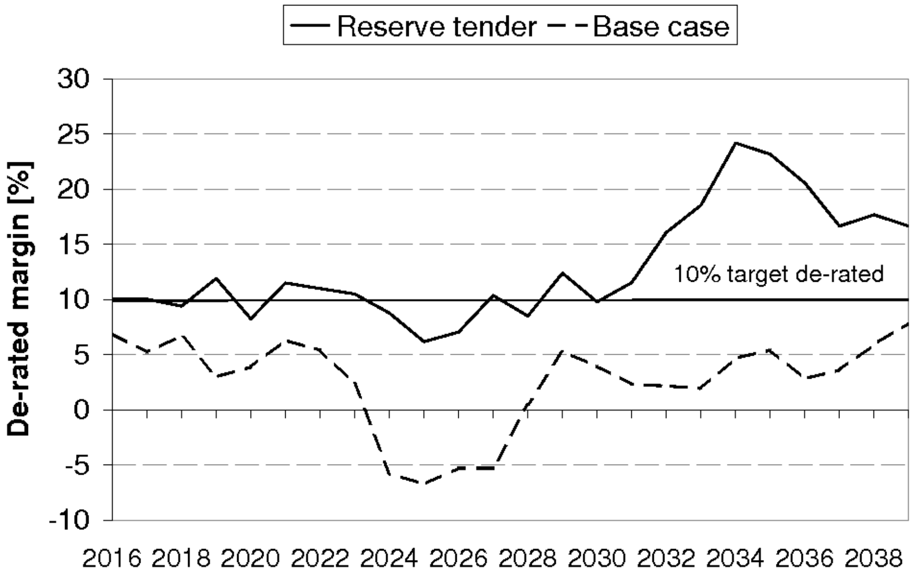


Figure 8.5: *Plot of de-rated capacity margins for strategic reserve against the base case during 2016-40.*

Fig. 8.6 compares the SRT tender OCGT capacity to the base case (i.e., in response to expected energy market and ancillary services revenues only) OCGT investment and mothballing. Also shown are the base case OCGT investments and mothballing (striped blocks). Introduction of the SRT was expected to prevent investment in OCGT capacity on account of tendered OCGT capacity damping expected market prices, and hence gross margins. Unexpectedly, the model results show that some OCGT capacity is still triggered as a result of expected energy market revenues, yet this is delayed by 2 years relative to the base case, and only occurs in 2017 and 2019 (Fig 7.17). This occurrence of investment in OCGT being driven by energy market revenues can perhaps be explained by the fact that, during 2017-19 high anticipated gross margins are still forecast due to the capacity retirements and demand growth after 2020. However as in the base case, energy market induced investment is not enough to meet the target de-rated margin. Consequently, large volumes of OCGT capacity are tendered in the run-up to 2025. In total, 28.1 GW of new build OCGT plant is tendered during 2013-25, with an average volume of 2.2 GW per year, although the variation in tendered volumes is remarkable. For instance, as little as 100 MW is tendered in 2014 with a massive 5.3 GW tendered in 2021, which indicates that procured volumes are quite volatile. Interestingly, after 2024 OCGT plant not included in the SRT begins to be mothballed. A small volume of capacity is de-mothballed in 2026, only to be mothballed again in 2027. In fact, by 2030, 6 GW of OCGT capacity has been mothballed. This can be attributed to de-rated margins climbing after 2030 (Fig. 8.5) and thus damping energy market prices, which has a negative impact on gross margins causing generators to mothball OCGT capacity. When generators take the decision to mothball, this results in the instantaneous removal of capacity. This explains why the SRT performs badly in 2025 (i.e., the target de-rated margin of 10% is not met). In the run-up to 2025, 850 MW and 2 GW of OCGT capacity is mothballed in 2024 and 2025, respectively (Fig. 8.6). This shortfall cannot be avoided by the SRT due to its 3 year lead time. In a more sophisticated model where OCGT generators participating in the energy market are able to forecast revenues from the SRT and, if more profitable effectively ‘switch markets’ from energy to SRT, then this OCGT mothballing may be prevented. Furthermore if such a mechanism was to be implemented, it would plainly be beneficial for it to include some form of incremental SRT less than 3 years out from the delivery year. The intention of these incremental tenders would be to incentivise generators who have mothballed capacity to bring it back on line if it is required. Note that no tendering occurred for the ten years after 2028 on account of high de-rated margins (as a result of new nuclear coming on line). This suggests that much of the tendered capacity will be utilised during the first years of its operation, but left idle for some years after that.

Inspection of the evolution of total installed capacity over time in Fig. 8.8 shows how installed capacity consists of a large volume of OCGT capacity relative to the base case (cf. Fig. 7.15). This shows that the response of the market to a SRT is to reduce CCGT investment (Fig 8.7) with all OCGT investment after 2019 carried out in response to the tender only (note that the positive purple bars after 2019 in Fig. 8.6 represent de-mothballing). More precisely, there is 33% less CCGT investment here compared with base case, in contrast nuclear investment remains largely unaffected. Consequently, by 2030 the generation market is dominated by non-dispatchable generation (e.g., wind and nuclear account for 43% of installed capacity, compared with 16% in 2010), with a small volume of mid-merit generation (e.g., CCGT accounts for 24% of installed capacity, compared with 38% at the start of the simulation), and over 32% (average of 28% over 2020-40) of the generation market as strategic reserve. Furthermore, these

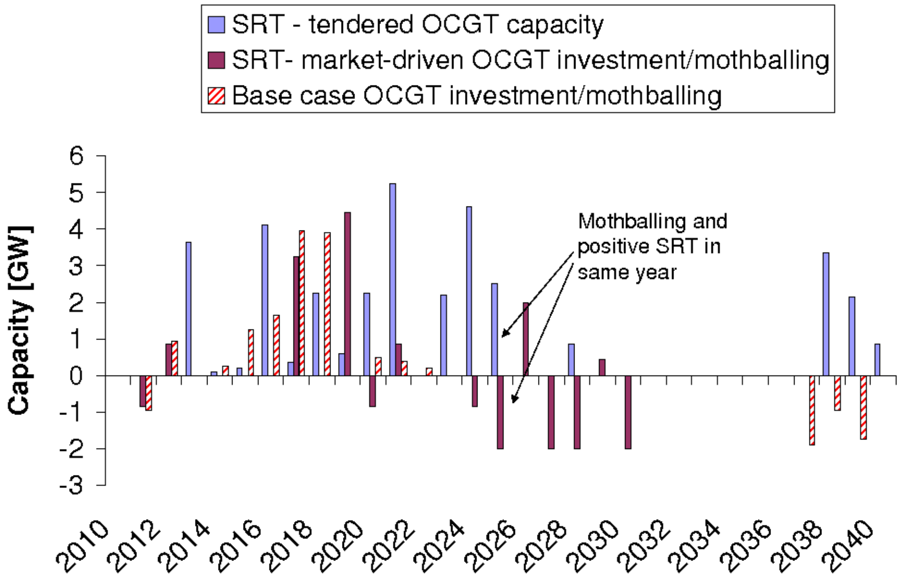


Figure 8.6: *Plot of OCGT investments and mothballing/de-mothballing under SRT versus base case during 2010-40. Note that 'market-driven' investment is in response to energy market revenues (i.e., the endogenous investment decision).*

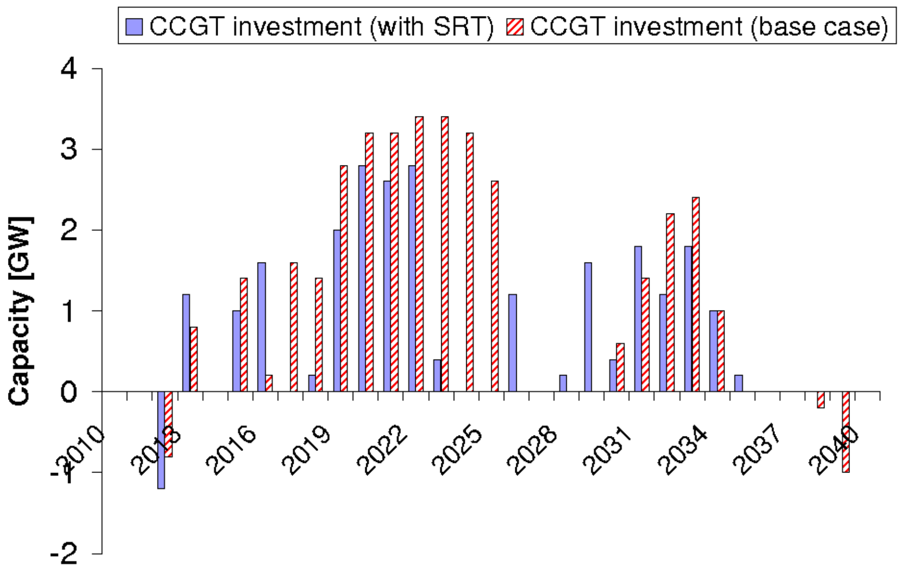


Figure 8.7: *Plot of CCGT investments and mothballing/de-mothballing under SRT versus base case during 2010-40.*

simulation results show that the SRT is required to meet shortfalls throughout the 2020s, though once large volumes of nuclear investments begin entering the system, de-rated margins climb above the 10% target (average de-rated margin 17.5% during 2030-40). Consequently, the system is holding a large volume of unused reserve in some years. From a policy perspective, it is arguably uneconomical to tender for such a large volume of peaking capacity in the lead up to, and thoughout the early part of, the 2020s, when it may only be required for the first years of its operation. That said, as shown in Fig. 8.6, positive tender volumes begin again in 2038 suggesting that they will be utilised again after a period of inactivity.

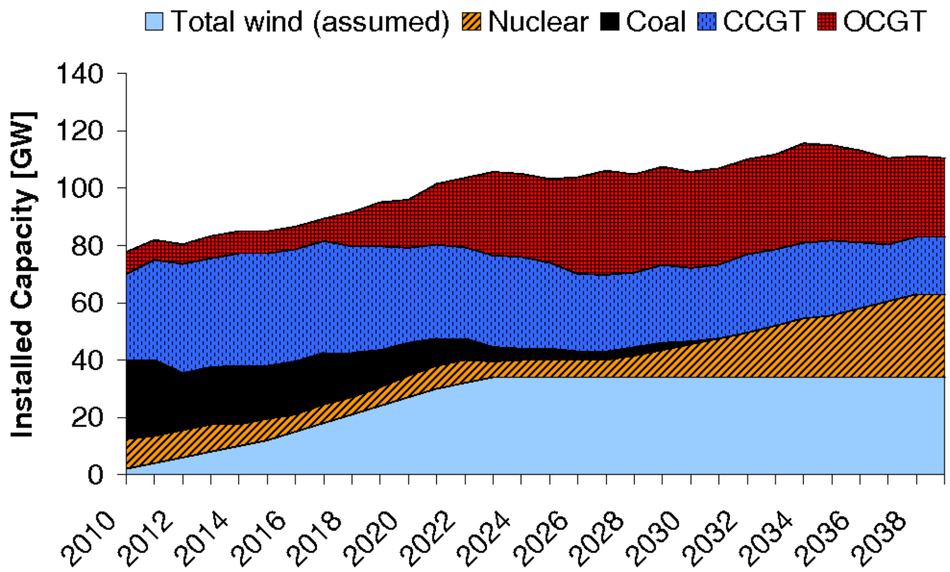


Figure 8.8: *Plot of total installed capacity over 2010-40 under the SRT, i.e., the result of the mix and amount of generation investment and retirements over time.*

8.2.4 Market-wide capacity market (after [54])

In this section a simplified version of the PJM's Reliability Pricing Model (RPM) is tested. The target of the RPM *"is to achieve the target level of reliability most of the time"* [51]. This target is based on loss-of-load probability 1 day in 10 years (or 0.0003) with procurement of capacity occurring 3 years before it is required. In this application, the target level of reliability is the 10% de-rated capacity margin.

There was a desire to keep the capacity market mechanism simple so that integration into the existing model framework could be achieved in a reasonable time frame. Consequently, many of the complexities of RPM (see [51]) are not included in this implementation. Firstly, no locational pricing is considered on account of the model being single bus. Secondly, deferral of capacity retirement, which is something that the RPM has been known to induce [51], is excluded on account of the fixed lifetime assumption, however altered mothballing patterns can be assessed. Thirdly, no demand-side resources or transmission projects are considered in addition to competing generating resources in the capacity market. Finally, only one auction (3 years out) is included here, while the full RPM includes "base-residual" (3 years out) and three "incremental" (20, 10 and 3 months out) auctions.

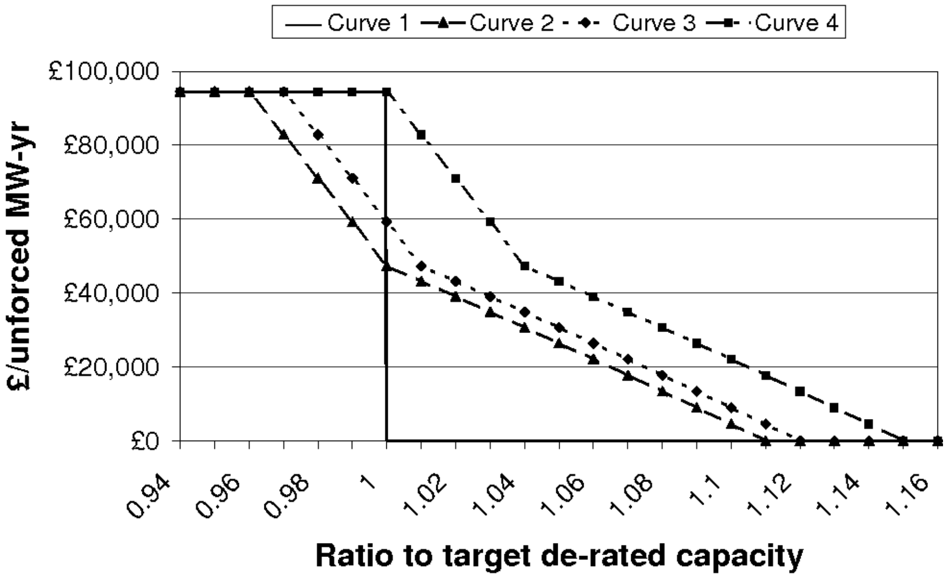


Figure 8.9: *Alternative capacity market demand curves used in this study. Inspired by [54].*

Like the RPM, a capacity demand curve is used to determine required capacity margins as a function of capacity prices. In addition to the classical vertical demand curve, a number of downward sloping demand curves are tested, these were inspired by [54]. These curves are known as Variable Resource Requirement (VRR) curves in the RPM [51] and are displayed in Fig. 8.9. It is important to note that these curves are a function of target de-rated capacity, not de-rated margin (one can be derived from the other because the forecast system peak load in (2.14) is known, e.g., Fig. 7.12). This method avoids negative ratios when the actual (projected) margin is negative and furthermore, these functions are consistent for different values of the target margin. The highest price for capacity (e.g., ratio of actual to target de-rated capacity below 1 in vertical demand curve) is based on twice the annualised FCs for a new entrant OCGT unit minus expected annual energy market and ancillary services (AS) revenues, known as the "net cost". The VRR curves used here are inspired by those tested in [54]; "Curve 1" is similar to the "no demand curve", "Curve 2" the "net cost at installed reserve margin (IRM) curve", where IRM is the ratio of installed capacity over peak demand [54], "Curve 3" the "net cost at IRM + 1% curve" and "Curve 4" the "net cost at IRM + 4% curve". The curves are calibrated so that the capacity price is precisely the net cost of the OCGT unit when the ratio of actual to target de-rated reserve capacity is 1 (curves 1 and 2), 1.01 (curve 3) or 1.04 (curve 4). Note that the curves tested in [54] included predicted energy market revenues when calculating annualised net costs; this was determined using the relationship of energy market and AS revenues to unforced (or de-rated) reserve for a hypothetical new turbine in a dynamic simulation model of the PJM market [54], and empirically using *"the actual values that would have been experienced for such a turbine in years 1999-2004"* [54]. Taking a similar approach, the relationship beween energy market gross margins from the 30 year base case simulation versus the target de-rated capacity required to meet a de-rated margin of 10% are analysed. These are displayed in Fig. 8.10. The points show the realised OCGT gross margins (GMs) during each year of the 30 year base case simulation. Only ratios below around 0.96 yield positive gross margins and furthermore, the average GM for the values in the range [0.94, 0.96] is -7.32 £/unforced MW-yr. As a result of this rather bleak picture for OCGT GMs, the maximum capacity market price is set at twice the full net cost of a new entrant OCGT plant (i.e., $2 \cdot 47,250 = 94,500$ £/unforced MW/yr, Table 7.5) and the VRR curves 1-4 in Fig. 8.9 have been calibrated according to the descriptions above and gradients described in Section V-A of [54].

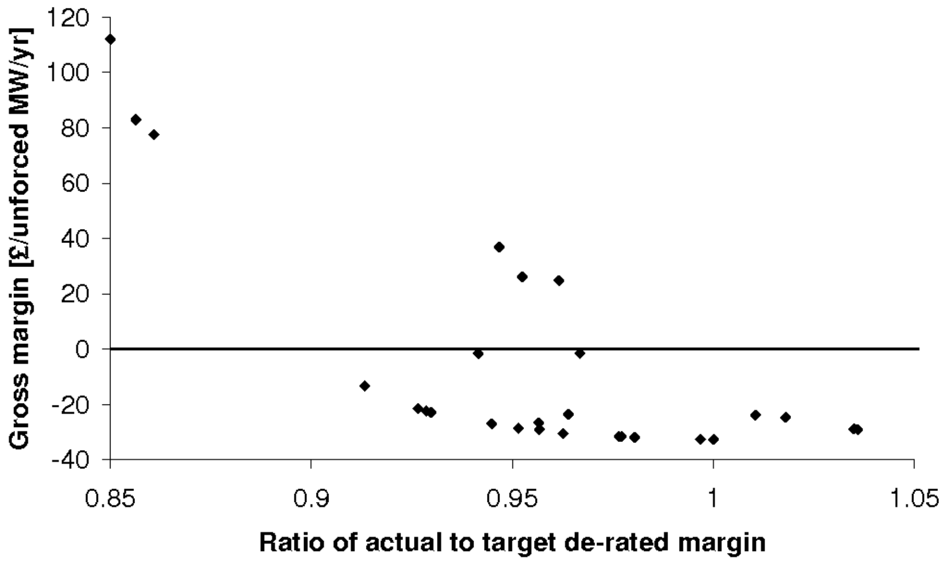


Figure 8.10: *Relationship of OCGT energy market and AS revenue gross margins to de-rated margin, expressed as a ratio to the target.*

8.2.4.1 The capacity market auction process

Capacity is offered into the auction by competing generators (or generators, demand and transmission in full RPM model) and an aggregate supply curve is calculated. The market clearing price is obtained from the intersection of the supply and the demand curves. An example of this is shown in Fig. 8.11; the graph shows the PJM capacity market auction results for the 2012/13 planning year at the 3 year ahead stage (i.e., the "base-residual" RPM capacity market auction). Note that, for the RPM auction, capacity is offered based on unforced capacity (UCAP), e.g., if a 50 MW unit has a FOR of 0.05, then $50\cdot(1-0.05) = 47.5$ MW is offered into the auction. To maintain consistency, this is referred to herein as de-rated capacity. Although included purely for demonstration purposes, the clearing price in Fig. 8.11 is expressed in $/MW-day, which can be translated into £/MW-yr (as in Fig. 8.9) by multiplying by 365 and using the appropriate currency conversion.

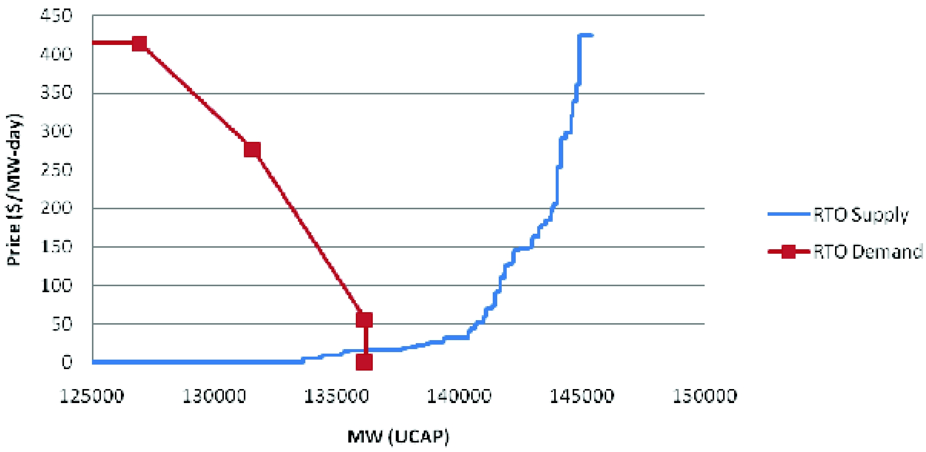


Figure 8.11: *Example PJM RPM clearing for 2012/13 for offered UCAP (supply) and VRR curve (demand). Intersection of Regional Transmission Organization (RTO) supply and demand curves determines the auction clearing price. Source [228] (used directly).*

In this implementation, the generator supply curve at each yearly auction is constructed using the total installed capacity (i.e., all capacity participates in the auction), which is de-rated and offered into the auction. For simplicity, this is modelled as a single vertical supply segment, and the capacity market clearing price, which is paid to all generation resources per de-rated MW, is the intersection between the total de-rated supply and demand curves. The investor includes the capacity price at each decision year and assumes the current price will be sustained for all

7 of the stochastically simulated years when calculating the value of an investment. Therefore (5.17) becomes:

$$V_x = \sum_{i=\tau_x}^{\alpha_x} \frac{GM_x^i - FC_x + CMP_x}{(1+r)^{i-\tau_x}} - (IC_x + DC_x), \tag{8.4}$$

where CMP_x is the capacity market price.

8.2.4.2 Simulation results

Fig. 8.12 shows the evolution of de-rated capacity margin over the period 2016-40 for each of the demand curves tested. The evolution of the mix and amount of generation over time for each demand curve is shown in Fig. 8.13. The graph shows how de-rated margins improve significantly relative to the base case. It should be noted that given the uncertainty in many of the input parameters used to calculate expected de-rated margins, it cannot be said with a high degree of confidence, that the simulated values of de-rated capacity margin will be realised. Therefore, once again, rather than focus on absolute values, this analysis focuses on the performance of the demand curves relative to one another and to the base case. Table 8.1 provides summary statistics on average de-rated margin and capacity prices (Fig. 8.9) for the 30 simulated future years.

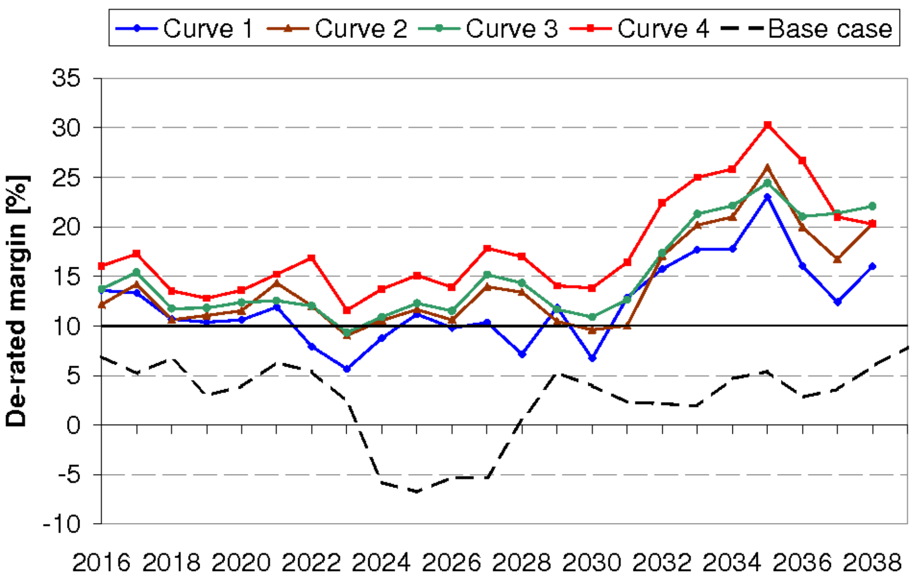


Figure 8.12: *Plot of de-rated capacity margins for the 4 capacity market demand curves tested relative to 'energy-only' market base case during 2016-40.*

	De-rated margin [%]		Capacity price [£/unforced MW-yr]		Ratio to target de-rated capacity		Years target met
	Average	SD	Average	SD	Average	SD	
Base case	5.6	7.1	-	-	-	-	5/30
Curve 1	11.9	3.8	39,679	43,144	1.02	0.03	21/30
Curve 2	13.9	4.2	36,307	27,784	1.04	0.03	27/30
Curve 3	14.7	4.3	38,191	29,419	1.04	0.04	29/30
Curve 4	17.1	4.9	41,933	31,019	1.06	0.04	30/30

Table 8.1: ***Summary statistics for market-wide capacity markets and the 'energy-only' base case.***

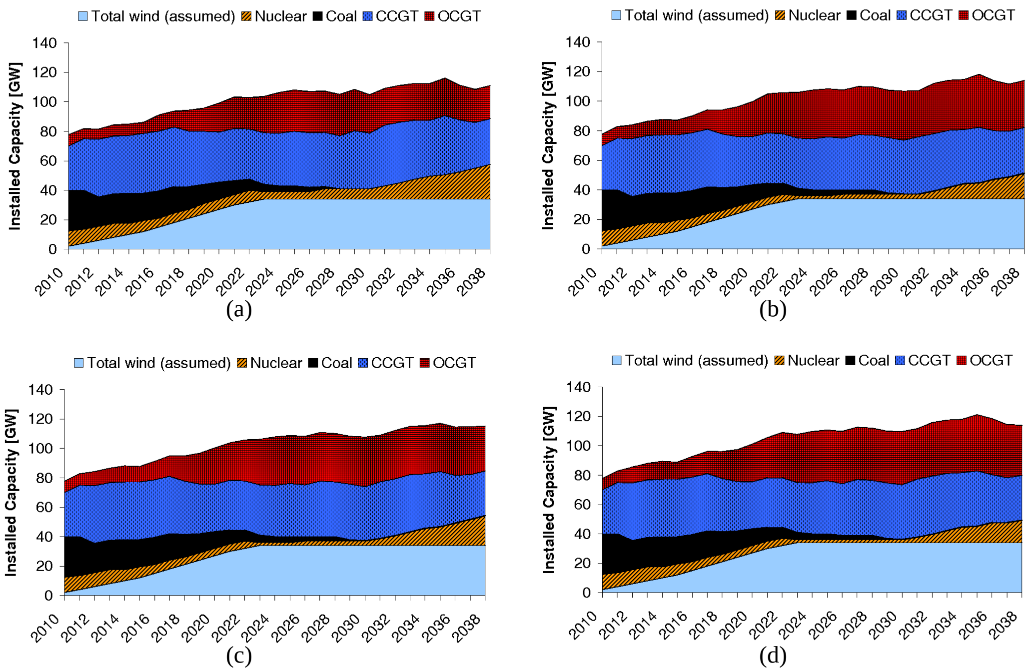


Figure 8.13: *Plot of total installed capacity over 2010-40 for the 4 capacity market demand curves tested (a) "Curve 1"; (b) "Curve 2"; (c) "Curve 3"; and (d) "Curve 4".*

The expectation would be that "Curve 1" would exhibit a very volatile capacity market price due to the curve paying either zero when above the target level of de-rated capacity, or twice the full net cost of a new entrant OCGT plant when below it. Further, as a result of this volatility, the simulation would display less investment and hence higher security of supply risk relative to the other demand curves, yet there ought to be an improvement relative to the base case. Results showed this to be the case, with the capacity price oscillating between zero and maximum throughout the simulation as Fig. 8.14 shows. Moreover, as shown in Table 8.1, the target de-rated margin is achieved in 21 of 30 years, compared with 5 of 30 in the base case. These findings are broadly similar to [54], where for a dynamic simulation model of the PJM market, the target de-rated margin was met in 54% of years compared with 34% with no capacity market.

The expectation would be that “Curve 2”, would be less volatile than “Curve 1”, again providing a better level of security of supply relative to the base case. This was indeed the case (Table 8.1), and the target margin is met in 27 of 30 simulated years with a much smoother capacity market price (Fig. 8.14). “Curve 3”, would be expected to be less volatile than “Curve 1”, again providing a better level of security of supply relative to the base case. However given that demand is willing to pay just 1% higher than “Curve 2”, the expectation would be that these curves would produce broadly similar results. This was indeed the case, with the reduction in price voliatility relative to “Curve 1” visible in Fig. 8.14. Here the target margin is met in 29/30 simulated years.

“Curve 4” would be anticipated to perform best in terms of security of supply risk because demand is willing to pay for capacity in years when available generation is expected (recall that the auction occurs 3 years ahead of delivery) to exceed the target level of de-rated capacity by up to 14% (Fig. 8.9). This is indeed the case, though the results suggest that this curve may be over-paying for capacity because de-rated margins are well above the target in most years. However given that all other curves experience at least one year when the target is not met, there is a higher degree of confidence that this demand curve mitigates the security of supply concerns present in the base case simulation.

Fig 8.14 shows the capacity market price for each demand curve, in £/unforced MW/yr, paid to each generator in addition to energy market revenues. There is clearly a boom and bust pattern to the capacity market in the early years, particularly under “Curve 1”. Recall that “Curve 1” is designed to pay either zero when de-rated capacity is equal to or above the target or twice the full net cost of a new entrant OCGT plant when below the target. This leads to a highly volatile price relative to the other curves which provide a smoother payment profile (Fig. 8.9). For curves 2 to 4, the capacity price is positive in all years out to 2030, although it reduces after 2030 in response to new nuclear plant entering the system (Fig. 8.15), which results in high de-rated margins (Fig. 8.12). Fig. 8.16 shows the impacts on OCGT gross

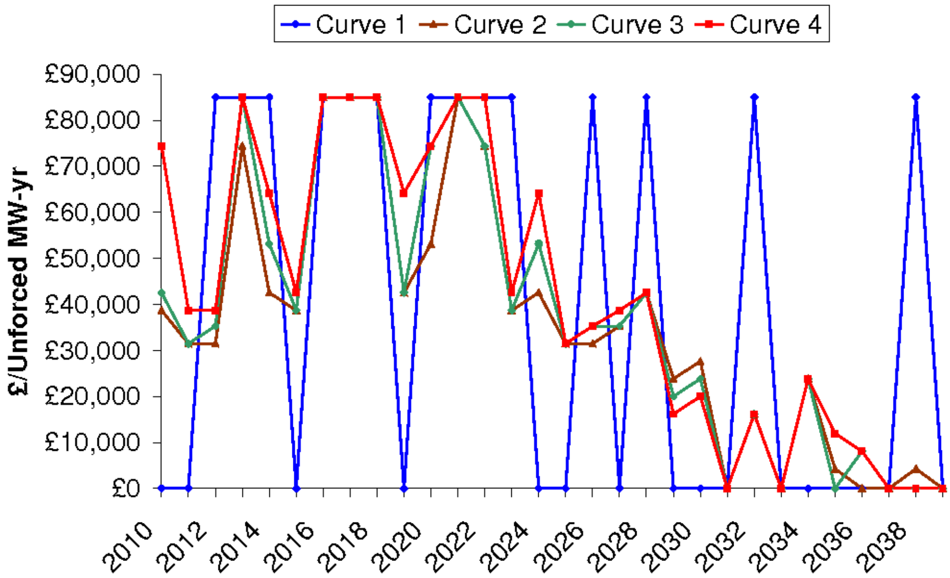


Figure 8.14: *Yearly capacity market prices for the 4 capacity market demand curves tested.*

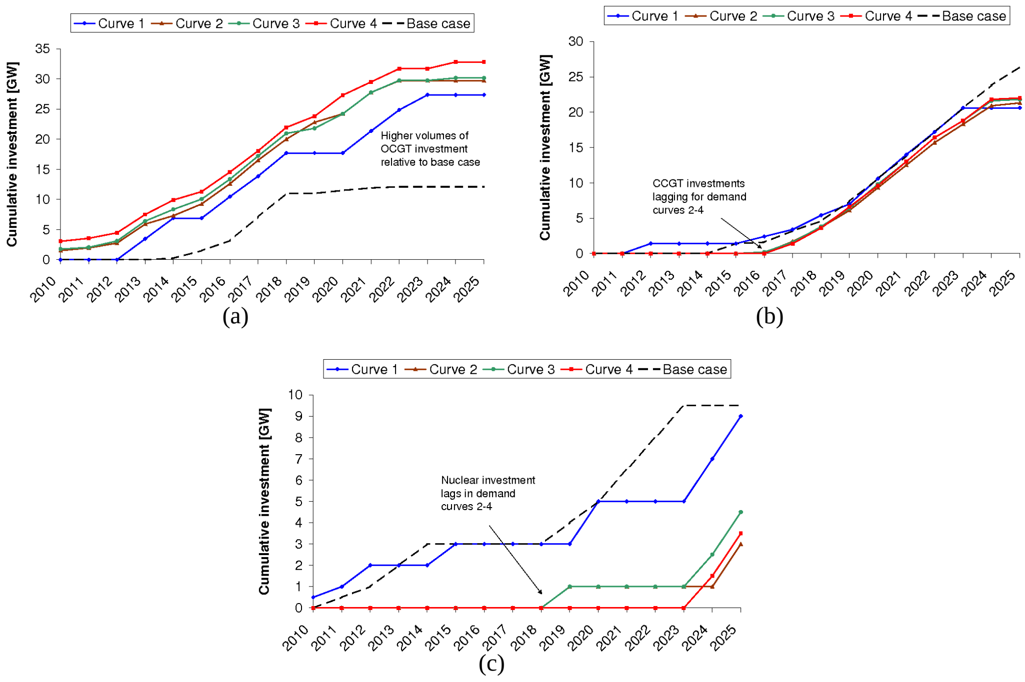


Figure 8.15: *Cumulative (a) OCGT; (b) CCGT; and (c) nuclear investment for the 4 capacity market demand curves tested relative to the 'energy-only' market base case.*

margins (GMs). Results for curves 2-4 show that OCGT generators receive positive and less variable GMs in most years out to 2030 (in fact, using "Curve 4", OCGT generators receive positive GMs in all years 2010-30). In contrast GMs reduce after 2030 and in some years and are in fact lower than for the base case (e.g., 2032). This occurs because more capacity is built under the 4 capacity market designs (Fig. 8.15) than in the base case. Consequently, generators receive less revenue from the energy market on account of higher levels of available resource (recall that price mark-up is reduces as capacity margins increase). Thus when the capacity price reduces, generator GMs perform similarly to under the 'energy-only' market simulation. A similar pattern emerges for CCGT generation; Fig 8.17, demonstrates how the variability of GMs is reduced relative to the base case, particularly for demand curves 2-4, though the outlook for overall levels of profitability across the 30 year simulation remains bleak. The average CCGT GMs were -37.8 £/kW, -33.6 £/kW, -25.26 £/kW -26.4 £/kW and -20.8 £/kW, for the base case, "Curve 1", "Curve 2", "Curve 3" and "Curve 4", respectively. This indicates that although profitability remains low, a relative increase across the 30 years was experienced for all demand curves tested. As it is relative difference that is of interest here, the conclusion must be that introducing a capacity market has a positive effect on CCGT and OCGT GMs, in addition to the level of generation adequacy risk.

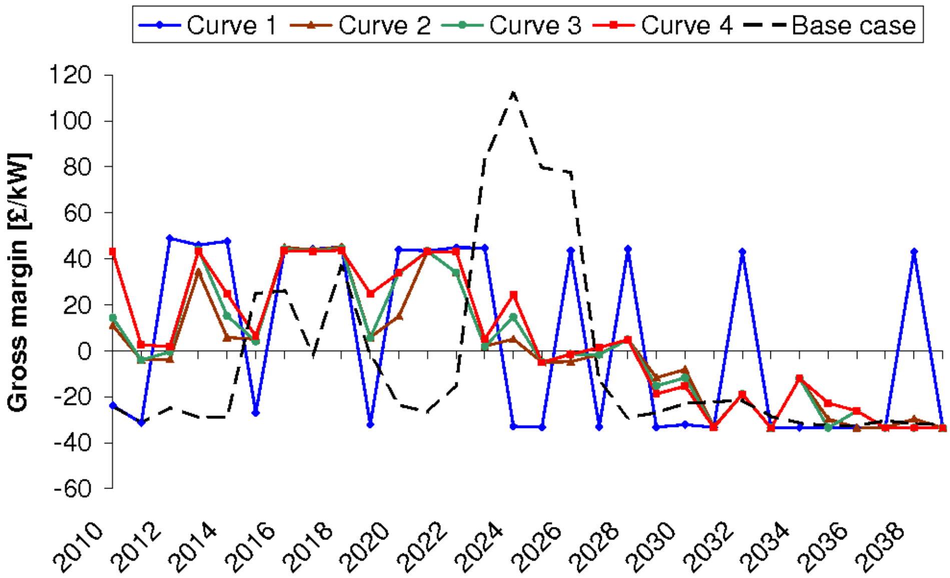


Figure 8.16: *Plot of simulated total gross margins for OCGT capacity for the 4 capacity market demand curves tested relative to the 'energy-only' market base case.*

Fig. 8.15 shows the cumulative trend in levels of investment relative to the base case. Note that this is the volume of investment triggered, not changes in installed capacity as in Fig. 8.13. An interesting observation when comparing "Curve 1" to "Curve 4" is the mix an amount of capacity over time in both cases. "Curve 4" provides a higher level of installed capacity and de-

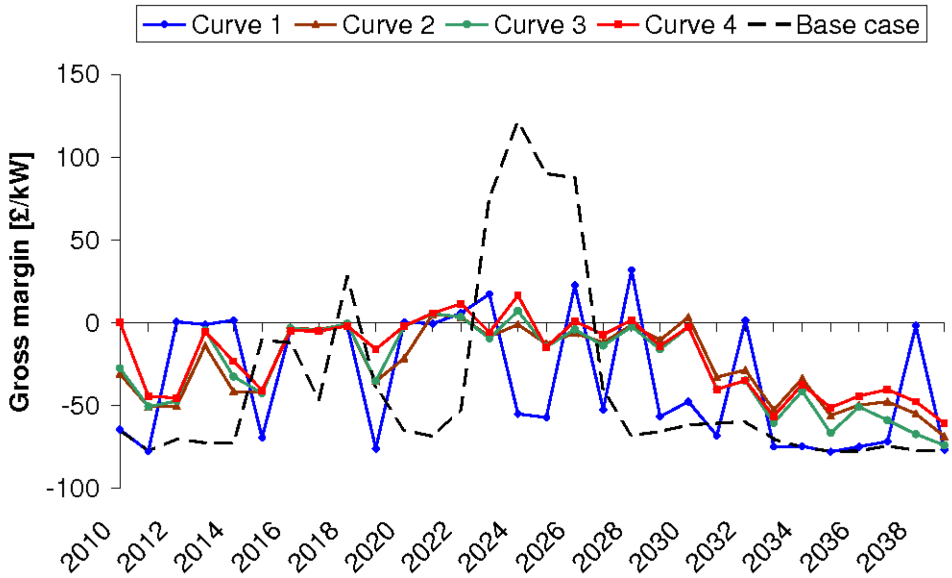


Figure 8.17: *Plot of simulated total gross margins for CCGT capacity for the 4 capacity market demand curves tested relative to the 'energy-only' market base case.*

rated margins, although the degree of diversity within the capacity mix reduces. For instance, nuclear capacity in 2040 is 21 GW and 15 GW for "Curve 1" and Curve 4", respectively. Moreover all four capacity market curves produce significant reductions in short and medium term nuclear investment relative to the base case (Fig. 7.15), and the capacity mix contains a much larger volume of CCGT and OCGT plant. Perhaps most striking, is the transformations in nuclear and OCGT investment; nuclear is significantly reduced early in the simulation (Fig. 8.15(c)), with OCGT investment much higher relative to the base case (Fig. 8.15(a)). This is attributed to the fact that the capacity market price is based on the total fixed costs of a new entrant OCGT, so it is hardly surprising that this type of generation becomes more attractive for investment.

A useful piece of analysis is to compare the total capacity market costs (i.e., total realised de-rated capacity multiplied by the capacity market price in each year of the simulation) with the reduction in unserved energy costs relative to the 'energy-only' market base case (i.e., no capacity market). This result would likely be of interest to policy makers seeking to answer the question: is a market-wide capacity market good value for money for consumers? Fig. 8.18 shows a plot of the cost of each demand curve tested per MWh of unserved energy reduced over 2016-40. The data suggest that "Curve 1" provides the least value for money (note the two

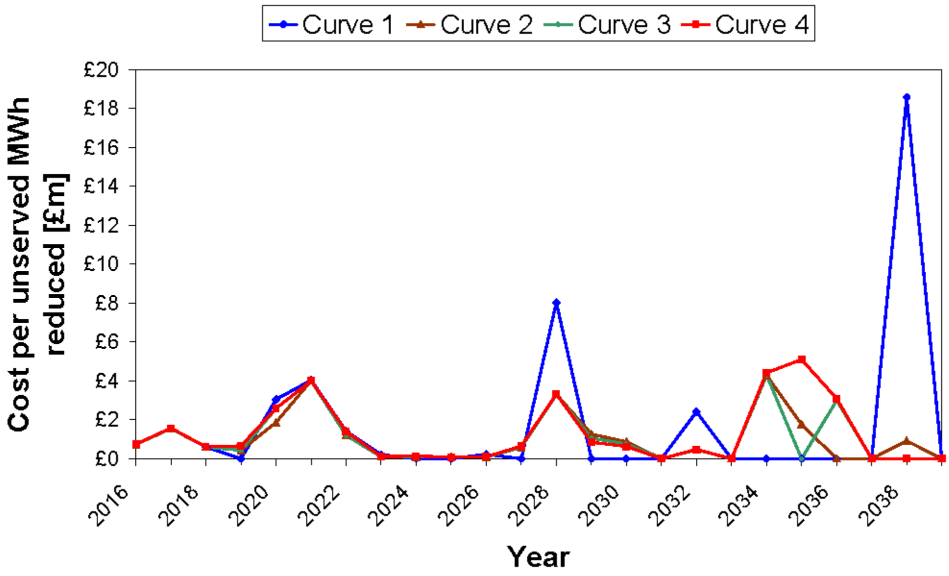


Figure 8.18: *Plot of cost of reduction (£million) per unserved MWh reduced 2016-40 for the 4 capacity market demand curves tested (Fig. 8.9).*

	Cost of unserved energy [£m/yr]		**Cost of reduction in** unserved energy (MWh) 2016-40 [£m/yr]	
	Average	SD	Average	SD
Base case	57.8	92.5	-	-
Curve 1	1.7	1.6	1.7	4.1
Curve 2	1.0	1.1	1.0	1.3
Curve 3	0.8	1.0	1.1	1.4
Curve 4	0.5	0.9	1.3	1.6

Table 8.2: ***Summary statistics for the cost of energy unserved in the 'energy only' market base case (cf. Section 7.5.1) and capacity markets, and the cost of the reduction in unserved energy (MWh) for the market-wide capacity markets over 2016-40.***

price spikes in 2028 and 2038), with all four curves experiencing high costs per unserved MWh reduced later in the simulation. The first 5 years of the simulation period are omitted from the plot as costs per unserved MWh reduced are an order of magnitude higher in some years (e.g., cost in 2013 are in excess of £25 million for all demand curves tested). This is due to the low levels of unserved energy across all cases (including the 'energy-only' market case) during 2010-15. The capacity market has positive prices in some of the first years of the simulation on account of the 3 year planning horizon (e.g., planning year 2013 forecasts de-rated margins for 2016, Fig. 8.14) . The lowest costs per unserved MWh reduced are experienced during the period 2023-27 for all the demand curves tested, this is hardly surprising given that this is the period of highest security of supply risk in the ‘energy-only’ market base case. Table 8.2 provides some summary statistics on the total costs of unserved energy for the ‘energy-only’ market base case and the capacity markets tested. Predictably the cost of unserved energy is significantly reduced with the introduction of a capacity market. For instance, average costs of unserved energy are 98% lower for “Curve 1” relative to the base case, with the other demand curves performing better still. Furthermore, “Curve 2” seems to offer the best value for money as the average cost per unserved MWh reduced is lowest at £1m, although this curve did not reach the target level of de-rated capacity margin in all simulated years (Table 8.1). Therefore a risk-averse policy maker may view “Curve 4” as the most attractive option even though its value for money appears lower relative to other designs. Generally speaking, this comparison suggests that the capacity market is worthwhile, although its value for money is lower in those years where generation adequacy risk is also low. The optimal policy may be to introduce the capacity market some years in advance of the period of highest security of supply risk (2023-28), though waiting too long may increase investor uncertainty and lead to an investment hiatus. Conversely, introducing a capacity market too early may lead to consumers over-paying for capacity. If used in an actual policy analysis, further tests would be required. For instance the mix and amount of capacity over time is different for each simulation (Fig. 7.15 and Fig. 8.13) so total year-on-year total system costs would also need to be compared. These costs include: 1) total production costs; 2) annualised fixed costs of the generation mix; 3) capacity market costs; 4) unserved energy costs; 5) system balancing costs; and 6) transmission reinforcement costs. This analysis is left for future research.

In summary, the introduction of a capacity market leads to a lower simulated level of generation adequacy risk, yet once the period of highest adequacy risk has passed, the capacity price reduces significantly. This has a negative impact on generator GMs, with CCGT and OCGT units once again unable to recover their fixed costs in most years after 2030. Further, the diversity of supply reduces: the capacity mix becomes dominated by CCGT, OCGT and wind generation during the 2020s with penetrations of nuclear generation not significant until after 2030. This can be viewed as a positive outcome: a high volume of flexible generation would complement the variability of wind.

These results demonstrate that introducing a capacity market mitigates simulated security of supply risk concerns during the 2020s and provides reasonably good value for money, yet the dynamic nature of the capacity market price means that generator revenue risk returns after 2030 when the capacity price reduces. The use of a mechanism that induces investment in the capacity required to meet a target level of security of supply risk (or a proxy for it) is sensible, although an analysis of costs per unserved MWh reduced shows that to achieve optimal value for money it may be desirable to allow the system to drop below the target level of de-rated margin in some years. The fact that these investments struggle in later years as de-rated margins recover is a less desirable outcome. It is interesting to note that the period of overshoot, 2032-37, is of a similar duration to the period of highest generation adequacy risk in the base case, 2023-28. This suggests that, depending on the generation adequacy policy in place, around 6-7 years of tight (or excessive) supply conditions will be experienced during (or after), a period of intense thermal generation investment in response to rapid wind capacity growth, thermal plant retirements, and demand growth.

8.3 Chapter summary

In this chapter the integration of two capacity mechanisms in the dynamic model has been presented. To begin, Section 8.1 provided an overview of the modelling approach, including why techniques from control theory could not be called upon to directly implement these mechanisms. Next, in Section 8.2, by drawing on the base case model results for the GB market case study, together with recent debate about generation adequacy in GB, the need for and design of a GB market capacity mechanisms was presented. This included a discussion on the socially optimal level of investment in sub-section 8.2.1. Results showed that the system may fall below the optimal level of reliability for a VOLL of 10,000 and 30,000 £/MWh. This was followed by a deeper analysis of the de-rated capacity margin results from the base case, in particular how this measure is dependent on the scaling factors applied to each generator. Following this, the implementation of the strategic reserve tender was described. Simulation results showed that the SRT is required to meet shortfalls throughout the 2020s, though once large volumes of nuclear investments begin entering the system, de-rated margins climb above the 10% de-rated margin target after which large volumes of reserve are unused in some years. This was then followed in sub-section 8.2.4 by implementation and results for the market-wide capacity mechanism. Results showed that the introduction of a capacity market leads to a lower simulated level of generation adequacy risk and, although less diverse relative to the base case, a supply mix that compliments the variability of wind generation. The results and interpretation provided here will be drawn upon in the next chapter where discussion and conclusions for the project are given.

Chapter 9

Discussion and Conclusions

This final chapter discusses the results of this research and provides conclusions. The thesis statement from Chapter 1 is answered and contribution to knowledge reviewed. Implications of thesis findings are discussed, including impacts on electricity market reform. Recommendations for further work and overall thesis thoughts and experiences are also shared.

9.1 General discussion of results

Like any project spanning a significant amount of time, the model implementation evolved as knowledge was acquired and understanding improved. There have been two main iterations of the model. In the first instance, the goal was to validate performance against the past. In light of the findings in the validation phase and feedback from a subsequent publication [194], the second goal was to develop the model to account for the high penetrations of wind power that are expected to emerge in GB.

9.1.1 The dynamic generation investment market model

Dynamic models can provide insights that are not always available when analysing a single moment or point of system balance. Further, the effect that the mix and amount of capacity over time has on the level of security of supply risk and generator profitability can be assessed.

In this thesis, a dynamic generation investment model, which captures the negative feedback mechanism that exists between market conditions, investor price expectations and the timing and intensity of year-on-year thermal capacity addition and withdrawal has been presented. In combination with exogenous model parameters which include forecasts for demand and wind capacity growth and production, the model has been applied to a GB market case study to forecast the capacity, price and reliability oscillations that could potentially occur.

9.1.1.1 Preliminary implementation

The initial task was to extend the work of Häni [123] and apply the model to the GB generation investment market. This was motivated by a key philosophy of control theory that the first task should be to model the behaviour of the real system. In this case, the real system was chosen to be the GB ‘energy-only’ electricity market since the introduction of the NETA framework in early 2001. The simulations against the historic market dynamics were scrutinised, particularly the replication of historic levels of capacity margin, and a number of limitations were discovered. For instance, the dynamic model took a simplistic approach to calculating investor gross margins. Furthermore, a robust methodology to investigate the impact of high levels of wind power investment on market dynamics was required. This was coupled with the need to estimate the expected contribution of generating units to serving load and the revenues they receive by doing so in a power system with a high wind penetration that is not necessarily in long-run equilibrium. As the simulation was to be run over many years (2010-40), this required a computationally fast, accurate and robust technique. An opportunity to undertake a 6 week placement with the Electricity Policy Research Group at the University of Cambridge was seen as an ideal opportunity to research such a method. During this placement, the MOND technique for approximating the LDC was investigated and was perceived as the ideal candidate to be developed and embedded into the dynamic capacity investment market model.

9.1.1.2 Final implementation

The dynamic investment model into which the MOND technique is embedded was updated and described. The model includes a Monte Carlo investment model which randomly sampled capacities, fuel prices and load growth. The distribution of potential profits are constructed from the MC simulations of the stochastic variables and a risk averse investor with Value at Risk decision criteria with $q = 5\%$ is assumed. New features of the investment decision were also developed, these included an aggregate investment response curve as well as a mothballing response curve based on the philosophies of Prospect Theory. The MOND technique was applied to calculate expected output, costs and revenues of thermal generation subject to varying load and random independent thermal outages in a market situation. It was adapted for use in the dynamic capacity market model with high penetrations of wind by performing a residual load calculation with simulated wind outputs. The MOND has proved very easy to convolve with plant outages for the expected output calculations. Further, expressions for calculating the expected revenues from price mark-up have been derived. The validity of these expressions was confirmed by Monte Carlo simulation, although as discussed in Section 7.2, if used for an actual policy analysis, the impact that the number of Monte Carlo runs has on the model outcomes should be evaluated and, if possible, larger sample sizes used.

9.1.2 Wind modelling

In light of preliminary model results, and observations about the relative success of policies outside of pure market rules in promoting investment in wind generation, it was decided to consider growth in wind capacity as exogenous to the model. This motivated an investigation into how best to incorporate production from increasing penetrations of wind within the dynamic investment simulation model. The techniques used to transform simulated and measured wind speed data for 2005-09 to power outputs have provided a contribution to knowledge in addition to the dynamic investment model work. Outputs from the production model were matched with empirical load data to estimate residual load duration curves for various penetrations of on and offshore wind. The results from this stage of the work were employed to update the investment market simulation model for an assumed level of installed wind, and estimate future thermal generation investment trends in GB.

Results showed dramatic changes to the shape of residual load distributions as the penetration of wind increases, with a trend towards normality at high penetrations. Further, the standard deviation of the distribution increases substantially with negative residual demand experienced at high penetrations. These negative loads occur for penetrations of 30 GW and above, yet the peak load remains largely unchanged. The residual load distributions provided an approximation of what thermal generation (in reality, also hydro power and other renewables) production requirements (ignoring network issues) would be for assumed future levels of installed wind capacity.

9.1.3 'Energy-only' market results

An 'energy-only' market setting with an existing installed capacity similar to that of GB has been used to estimate the economic profitability of capacity investments. Using relative levels of de-rated capacity margin and LOLE as the risk metric, simulation results for GB show that levels of generation investment lead to a increase in generation adequacy risk in some years, with erosion of de-rated capacity margins in the mid 2020s, and very tight supply conditions are experienced during a small number of peak hours. Many new investments, particularly peaking units, were unable to recover their fixed costs.

The relative increase in levels of security of supply risk was explained by inspection of the residual load histograms. The shape of the right-most tail suggests that even with very high penetrations, wind power does not contribute in all high demands periods. On the basis of the period analysed, the frequency of these high-demand/low-wind periods is too low to justify investment by private investors and it is these very high-demand hours when the potential for a capacity shortfall is highest.

9.1.4 Sensitivity analysis

A sensitivity analysis demonstrated that assumptions about investor risk profiles and expectations about new builds, load growth and the ability to exercise market power have a strong impact on simulated investment dynamics and subsequent levels of generation adequacy risk.

Overall, the model's qualitative behaviour was reasonable for the sensitivities tested and provided some useful insights, particularly when comparing the base case to the perfectly competitive market results. In all cases a pattern of increased relative levels of risk and erosion of de-rated capacity margins was experienced during the 2020s to some degree. Furthermore, the period of highest security of supply risk is 2023-28, though the magnitude of this risk differs between experiments. A prolonged period of increased security of supply risk is experienced throughout the 2020s for the perfectly competitive market case. Also providing investors with perfect foresight about capacity under construction produces less investment. This leads to more frequent periods of relatively high LOLE and low de-rated margins after 2020. In contrast generators experience positive gross margins in more years due to lower surplus margins, and hence higher prices. A sensitivity analysis on the primary fuel prices showed that lower coal prices (or indeed higher gas prices) may lead to different technology choice, with relative levels of high generation adequacy risk shifting to earlier in the time horizon.

9.1.5 Capacity mechanisms

The topic of the last stage of the research was to determine whether explicit capacity mechanisms such as tendering for strategic reserve (e.g., [67]) and capacity markets (e.g., [54]) can be designed to alleviate resource shortfall, prevent investment overshoot and achieve a 10% de-rated capacity margin. The results of the GB 'energy-only' market simulation indicated that such a mechanism may be desirable to improve reserve margins in the mid 2020s.

Results of the strategic reserve tender showed that the target is met in most years. The generation investment response to this mechanism is to reduce CCGT investment relative to the base case, with all OCGT investment after 2019 carried out in response to the tender only. Consequently, the generation market becomes dominated by generation outside the energy market, with over 32% of the generation market as strategic reserve. Such a high level of strategic reserve is arguably inappropriate in a liberalised electricity market. Moreover, a situation where the majority of generation investment is carried out in response to a centrally managed market undermines the credibility of the mechanism. Furthermore, the SRT is required to meet shortfalls throughout the 2020s, however once large volumes of new nuclear begin entering the system during 2032-37, de-rated margins rise well above the target. Consequently, the system is holding a large volume of unused reserve in these years.

Results of the market-wide capacity market demonstrated that introducing a capacity market mitigates simulated security of supply risk concerns during the 2020s, although the level of capacity market price volatility, and hence revenue risk, differs between the capacity market demand curves tested. The sloped demand curves perform best in terms of 1) meeting the target and 2) making generator revenue streams more predictable, yet some investment overshoot is experienced during 2030-35 on account of longer lead time plant entering the system. Consequently, generator revenue risk returns during this period. Furthermore, a comparison of the total capacity market costs to the reduction in unserved energy costs relative to the 'energy-only' market case suggest that on order to achieve optimal value for money it may be desirable to allow the system to drop below the target level of 10% de-rated margin in some years. However, it may be better to err on the side of caution and implement a more expensive capacity mechanism that delivers the target level of de-rated capacity margin in all years. These result would likely be of interest to policy makers seeking to answer the question: is the improvement in the level of security of supply risk delivered by a capacity market relative to 1) an 'energy-only' market; and 2) alternative capacity market designs good value for money for consumers? However, if used in an actual policy analysis, further tests would be required. For instance, the mix and amount of capacity over time is different for each realised investment scenario the 'energy-only' market and capacity market cases, so total year-on-year total system costs would also need to be compared. These costs include: 1) total production costs; 2) annualised fixed costs of the generation mix; 3) capacity market costs; 4) unserved energy costs; 5) system balancing costs; and 6) transmission reinforcement costs. This analysis is left for future research, perhaps integrated into the extensions suggested in Section 9.3.3.

In light of these findings, a more appropriate approach might be to encourage demand-side participation through smart grids and smart metering. Given that a significant numbers of smart meters are expected to be installed in GB by 2020 (e.g., DECC's goal of 53 million smart meters in 30 million homes and businesses across GB by 2020 [229]), this approach could be the most economical. That said, there is huge uncertainty surrounding the achievability of this target, and given that policy uncertainty delays investment, relying on smart meter delivery is risky.

If the model presented here were to be extended to include demand-side participation, then methods of implementation would have to be devised. Options include:

- Modelling demand-side participation as an endogenous model parameter. Including an additional generator(s) that could represent a volume of capacity (or energy) available for "dispatch", hence providing demand-side price response (DSR). The expectation would be that this DSR would be called upon at periods of peak demand when prices are typically highest (presumably the opportunity cost of a DSR consumer that has decided to reduce its demand would be lower than the VOLL, yet above the variable operating cost of the most expensive peaking unit). This DSR generator could be included in the probabilistic production costing and the effect of this additional capacity on expected output, costs and revenue of thermal generation estimated. The expectation would be that increasing the amount of available resource with a dispatch price below the VOLL would have a damping effect on prices and, if investors accounted for this DSR when forming their expectations, lead to less investment relative to the base case presented in Section 7.5. However less investment does not necessarily mean a less reliable system as the DSR capacity would lower the total level of de-rated capacity required at system peak.

- Modelling demand-side participation as an exogenous model parameter. If demand-side participation was driven by exogenous model parameters such as the availability of renewable resources, then profiles for demand net of wind production (and any other renewable resource for that matter) and demand-side response could be derived outside the model. These would then be provided as inputs for the MOND technique. This approach would also encapsulate the so-called 'smart grids' phenomenon, whereby at times of high demand and/or low available wind resource non-essential loads, such as certain household appliances or industrial processes, could be switched off or have their usage delayed until a more convenient time. Conversely, at times of high (and possibly surplus) production such processes could be enabled [230]. This method could also be applied if an energy storage device were being considered. These profiles would take into account the timing (e.g., base or peak demand hours) and level of demand reduction expected. Note that this method of implementation will not explicitly capture the feedback between prices and demand-side participation (as described in the previous bullet), so the "decision" to reduce demand is based on factors outside of pure market rules. However, this could be captured implicitly. For example, a group of consumers connected to a smart meter who increase demand during high wind hours (wholesale prices likely to be lower), and reduce it during low wind hours (wholesale prices likely to be higher).

9.1.6 Reliability of results

When making projections about the future levels and timings of generation investment, it is important to remember that estimates about profitability are inherently linked, and therefore sensitive, to the price of raw materials, exchange rates and the state of financial markets. Furthermore, economic projections are very difficult to make given the huge uncertainty surrounding estimates of market structure, technological progress, investment decision processes and policy. Unforeseen time delays and technical issues, such as lower than expected availability makes estimation problematic. Other confounding factors include the time scales required for investment appraisal, not to mention licensing and planning. That said, economic projection modelling can provide some insights providing uncertainly is addressed and results interpreted accordingly.

As noted in [20], a simulation model such as this shows *"the occurrence of cycles based on decision rules assumed by the model, where the behavior of the investors is still an empirical question"*. For instance, the degree of heterogeneous investment response is not modelled directly, however the investment response of the model is a smooth function of prices and costs, reflecting how a heterogeneous group of market participants would likely respond to changes in market conditions. Moreover, unlike other simulation models of a similar category, this implementation has been applied to an existing market, which included a simulation of historic investment trends. The focus has been on the likely behaviour of merchant generation investment in market environments. Given that the simulated investments and levels of capacity margin were not unreasonably different from reality in both the preliminary and updated model, the reliability of the results can be interpreted with a reasonable, but not high, degree of reliability. For instance, the level and timing of CCGT investments in both the historic simulations, showed 1) that the technology of choice was as in reality at that time and 2) that modelling a single investor to represent the aggregate response of the market was reasonable approach. If the model presented here were to be used for an actual policy analysis, a number of key areas of future research are recommended. These include: 1) a survey of investors regarding their use of analytical approaches when evaluating investment decisions under uncertainty and, if appropriate, modification of investor risk-averse decision making, 2) to develop an explicit representation of the heterogeneity of investor response, and 3) modelling endogenous bilateral contracting.

That said, no economic projection model should ever be viewed as able to mimic the behaviour of market participants precisely, and estimates of key inputs, such as the price mark-up function, have a large impact on simulation results. Results from simulation models that look to the future, which is hugely uncertain, must be viewed with a degree of realism and caution. More precisely, the absolute values of, for example, capacity investment and realised levels of de-rated margins and LOLE projected in this thesis are very much dependent on the model input parameters (e.g., assumed investor risk profiles and their use of analytical approaches when evaluating investment decision under uncertainty, FORs and residual load distributions). A more robust method of analysis is to look at relative levels of change to the sensitivities and capacity mechanisms tested. For instance, the introduction of a both SRT and a market-wide capacity mechanism certainly increased projected de-rated margins in the model, however to say that if implemented, both mechanisms would achieve the 10% target in all years would be unwise and further research is required.

One of the key factors limiting the reliability of the absolute values output by the model, is the limited availability of data on the generation side of the GB market. This is a commonly held view across research groups undertaking research in the power sector. For instance, [231] has recently called for improved transparency in the GB market and a key reason for change was that *"academic and technical research into emerging problems is all but impossible"*. Further, [1] states that *"investors in new generation capacity, particularly small market players, need access to market information, a well-established marketplace and regulated access to this marketplace."*, suggesting that the issue is not restricted to the academic community. The conclusion here must be that the degree of competition would benefit from a higher degree of transparency.

Other limitations of the modelling approach include:

- This implementation of the MOND technique does not consider the possibility of available wind generation exceeding either 1) available export capacity in a generation pocket due to transmission congestion or 2) raw demand net of inflexible base load (e.g., nuclear). Considering each in turn, firstly, if a representation of the load and available wind production in each region of the network is available, then curtailment of wind production due to transmission congestion could in theory be assessed by multi-area production costing methods. Secondly, future applications of this method could approximate the effect of inflexible generation by dispatching its must run capacity first and assuming a zero or negative price for the portion of the time that this capacity is on the margin.

- The load duration curve method (of which the MOND is a particular case) does not consider the possibility that ramp rate limitations could also result in wind spill.

- The residual LDC approach removes the chronological issues that arise in the wind and load time series. This is particularly important in the presence of large amounts of hydro and pumped hydro generation where chronological production costing methods may be preferred to load duration curve methods. However this implementation of the MOND technique is applied to the GB power system where the amount of hydro and pumped hydro is relatively small (about 4% of capacity), so the use of a load duration curve approach is reasonable.

- The impact of demand-side participation and smart grids on investment dynamics and realised levels of security of supply risk. This is left for a topic of future research, with possible methods of implementation discussed above in Section 9.1.5.

- Although the dynamic model has a stochastic element when looking forward and undertaking investment decisions, the ‘real-time” model is deterministic. As a result, parameters such as realised peak demand (99.9% load distribution percentile), generator FORs (expected values), demand growth and fuel prices (following DECC projections) are all based on one reality. To bestow more confidence on the model projections, a model where realised values are also stochastically simulated is required. However given that the 30 year simulation execution takes between 525 and 1575 minutes, adding this level of complexity is likely to make execution time unworkable, although increasing parallelisation of computing may make it possible.

- By taking an aggregate approach to modelling, the granularity of the aggregate supply curve is reduced. This aggregation was necessary in a model such as this where investment over a long-term time frame was under investigation. A more sophisticated model where the ability of new entrants or existing participants to gain “better than average” scarcity rents compared with other generators in their technology category by, e.g., having superior thermal efficiency, may provide more robust estimations of realised and expected future wholesale prices. Equally, as generating units age, they tend to experience a reduction in thermal efficiency and operating ranges on account of component preservation. One approach would be to partition technology types into sub-categories based on age, and assign thermal efficiencies accordingly (the older the unit, the lower the thermal efficiency). This also requires estimates of these efficiencies, and would increase the complexity and running time of the model, perhaps without gaining significant additional insights.

- Analysing only five years (2005-09) of empirical load and simulated wind data is arguably not sufficient to gain a robust statistical representation of the wind resource and its relationship with demand. Furthermore, this approach assumes that historic trends provide a fair reflection of the future. This is a heroic assumption given the changes in climate, and hence weather patterns, that are expected over the coming decades. This remains the best available method of prediction.

- The assumptions for plant availability used here are also up for debate. For instance, new nuclear builds were assumed to have 90% availability, yet the availabiltiy of the GB nuclear fleet has been well below this level and has varied considerably over time; from 80% in 1990 to 49% in 2008 [232]. Of course, these units are aging so longer and more frequent periods of scheduled outage are to be expected, but these figures demonstrate the variability of annual of plant availability (see [71] for impact on underlying risk calculations), and why assuming high availabilty for the entire plant lifetime is perhaps an unrealistic assumption to make.

- In the GB case study application, neither of the capacity mechanisms tested considered the interaction between capacity markets and the GB STOR market. This would need careful consideration if a mechanism of this type were to be implemented in GB. Details such as the fact that the PJM market ISO, on which the market-wide capacity market tested here is based, is a not for profit organisation and the GB market SO is a for profit organisation with incentives to reduce total costs of system operation, would also need to be considered.

9.2 Implications

9.2.1 Modelling developments

Under a liberalised market framework, electricity generating capacity investment decisions are taken and (in the most part) financed by privately owned generating firms. More precisely, firms participating in the ‘energy-only’ market undertake new builds or mothball existing plants in response to wholesale market price signals. It is precisely this price feedback mechanism that has been the focus of this thesis.

In light of this feedback concept, the modelling approach presented demonstrates that techniques from modern control theory can be employed to model the dynamics of generation investment in liberalised electricity markets over a long-term time frame. Although system stability analysis is not feasible, integration in the dynamic model of capacity mechanisms has also been achieved using more traditional methods of implementation, i.e., setting a target de-rated margin and introducing additional generator revenue streams based on observations of the expected realised de-rated margin 3 years ahead.

Further, the mixture of Normals approximation technique for load duration curves has been extended and embedded within the dynamic model, thus providing a robust methodology for future models of this type. Moreover, methods to calculate expected revenue from price mark-ups due to market power, a justifiable representation of oligopolistic power markets, have been derived here for the first time.

9.2.2 Generation investment

Recalling the thesis statement given in Section 1.3, the results of the GB case study suggest that the mix and amount of thermal investment in an ‘energy-only’ market with a high wind penetration will lead to increased relative levels of security of supply risk during the 2020s, with the tightest supply conditions coming during 2023-28. Further, simulations show a noticeable increase in security of supply risk under both the perfectly competitive (i.e., no price mark-up) and imperfect market experiments. This was despite intense investment in thermal capacity, in particular CCGT, and rapid exogenous growth in wind capacity. Analysis of generator gross margins showed that many of the investment undertaken out to 2023, although receiving good returns during the first years of operation when adequacy risk is highest, a bust phase develops after 2026 when large volumes of new nuclear capacity begin entering the system and security of supply risk, and hence revenue, is dampened.

Given that the model assumes the representative investor is capable of securing the necessary debt to finance large-scale capacity investment and in reality *"utilities are under pressure to moderate or lower their capital expenditure programmes and to find higher-yielding and higher-growth opportunities as a result of high debt levels, pressure to grow dividends, falling share prices and increased pressure on credit ratings"* [67], the intensity of investment may in fact have been over-estimated here. On the other hand, the model assumes a fixed lifetime for all capacity and an alternative outcome might be that generators, particularly in mid-merit units such as CCGT, delay investment in new capacity by switching their units to open cycle, thus increasing their SRMC, and running at lower utilisation factors, rather than risk building a new mid-merit unit that may not provide adequate returns on the investment.

9.2.3 GB electricity market reform

In the coming years, the GB power sector is set to undergo arguably one its most significant transformations to date. Underpinning this change is the need to mitigate the impacts of human induced climate change by reducing GHG emissions from the power sector. There is a pressing need to make the transition from a installed generating capacity which is dominated by fossil-fuel generation to one with high volumes of zero and low carbon generation. Already a number of bold environmental targets are in place, for example, the goal of supplying 15% of energy from renewable supplies (equal to 30-35% renewable electricity) by 2020, and an 80% reduction in overall GHG emissions from 1990 levels by 2050. Further, legislation has been put in place to limit, and ultimately withdraw, large volumes of GHG-intensive fossil-fuel generation. For instance, a significant volume of capacity is due to retire under the Large Combustion Plant Directive in the run-up to 2016. Under this legislation, over 11 GW of fossil fuel plant ($\sim$8 GW coal and $\sim$3 GW oil) is due to shut down. According to the UK Government, it is likely that around 30-35 GW of new electricity generation capacity will be required over the next two decades and around two thirds of this capacity will be needed by 2020 [233]. In addition, the GB nuclear fleet is reaching the end of its design lifetime with all but 3.5 of the existing 10.5 GW due to retire before 2020. This is a significant loss to the system and new capacity must be built in order to replace these retirements.

As already discussed, the UK Government is developing policies whose combined goals are 1) to promote the level of investment in renewable generation required to meet these targets;

2) maintain the level of security of supply risk that society has come to expect; and 3) reduce customer costs. These goals have introduced uncertainties that the current BETTA market framework is struggling to handle efficiently. The view of Government is that these goals will not all be met by the current market, and consequently a reform of market trading arrangements has been called for.

The results presented in this thesis can inform the debate about market reform, in particular the response and profitability of thermal investments to policies promoting investment in renewable technologies. As this model assumes that installed wind capacity will grow to a level likely to be adequate to meet the 30-35% renewable electricity target by 2020, the results here focus on the second goal outlined above, i.e., the level of security of supply risk. Simulation results have shown that if the market remains as an ‘energy-only’ oligopoly market, the level of investment required to offset retirements to existing capacity and to meet assumed demand growth, will not be adequate to allow the system to maintain the target level of security of supply, i.e., achieving a winter peak target de-rated margin of 10%. That said, the low frequency of the tight supply conditions that result from inadequate generation being built, suggest that the system may be able to cope without additional capacity mechanisms. Instead, targeting the demand-side through demand-side response and smart metering may be a more efficient method of avoiding power shortages. More interconnection with Europe will also alleviate some security of supply concerns, which seems a natural solution given that a number of EU directives on gas and electricity are aiming to open markets in other EU countries up to competition [234]. However if a capacity mechanism is the prefered course of action, then the results from the final stage of this work suggest that both a strategic reserve tender and capacity market will reduce security of supply concerns, though a degree of investment overshoot and generator revenue risk after the period of highest adequacy risk is anticipated.

9.3 Recommendations for further work

9.3.1 Multiple agent model

A topic of subsequent research could be the integration into the dynamic investment model investors with different attitudes to risk and methods of evaluating investment decisions under uncertainty. In this implementation alternate perceptions of reality and attitudes toward risk were tested during the sensitivity analysis, by comparison multiple agents competing together in a single simulation was not included. Further with 6 firms dominating the supply market in GB, an interesting and potentially valuable piece of analysis would be to represent these competing firms in the model. It could then be assessed whether a desire to maintain market share in the face of plant retirement and demand growth leads to similar investment dynamics presented here, or whether firms account for the effect of their investment upon prices and mark-ups for their existing fleet and deliberately do not invest in order to keep prices high.

9.3.2 Complimenting the stochasticity of wind production

The estimates of wind production presented in Chapter 6 could form a basis for an investigation into the need for, and design of, generation technologies best suited to compliment the variability and stochasticity of high penetrations of wind. This includes the technical and operational characteristics of these complimentary sources of generation. In this thesis, gas-fired plant such as OCGTs are the primary source of flexible output power, although in a focused analysis, options for energy storage will need to be considered.

9.3.3 Transmission network

The model presented here does not consider the impact of the transmission network on investment dynamics. Therefore, the next stage of investigation could be to assess the impacts of transmission on generation investment. An area of possible future research would be to extend the MOND technique to a two-area (e.g., Scotland-England) system with a transmission constraint. This could then be used to assess the impact of congestion on wind power availability and thermal generation investment in each area. This could be implemented using a Mix of Bivariate Normals reflecting the correlations in loads and wind in each area (e.g., as in [210]), based on two area production costing methods (e.g., see [235]). To maintain computational efficiency, the model could draw upon scenario reduction schemes (e.g., see Chapter 7 of [236]) when applying multi-area production costing and making an assessment of curtailment of wind production due to transmission congestion.

Furthermore, in GB, the *transmission network use of system charges* reflect the GB SO's preferred locations for new build generation. These vary across the system and can be negative, typically in areas close to high load centers, or positive, typically for remote locations. With a review of these charges currently underway via OFGEM's *Project TransmiT*, the impact on the location and volume of investment over time of alternative charging arrangements can be assessed. For instance, "real-time" locational system marginal pricing is being considered. Under this framework, each node on the network, would have a different marginal price for power. Advocates of this approach state that it reduces the ability of generators to exercise market power in a constrained network (be that offers to increase, or bids to reduce production). An interesting piece of analysis would be to investigate the impact of these reductions in generator surplus (or short-run profit) on investment and security of supply risk dynamics. This could be represented in the model by reduction (or even removal) of price mark-up in some or all areas of the network (which would have to be represented in the model). This would inform the debate about whether reductions in generator surplus under locational pricing leads to less investment or just the right amount in the right locations.

Once the transmission network is accounted for, the impacts of congestion on a market-wide capacity mechanism can be assessed. For instance, a mechanism that does not consider congestion when procuring for capacity may lead to undesirable outcomes such as a situation where the SO is not able to utilise capacity when it is needed due to the plant being located behind an active transmission constraint.

9.3.4 Electrification of heat and transport

In this application of the model, the use of electrical energy in GB is assumed to continue into the future. If the UK is to achieve its ambitious GHG emission reduction targets, then electrification of both the heat and surface transport sectors is required. For instance, the CCC have stated that a *"44% emission reduction in surface transport can be achieved by 2030"* (relative to 2008 levels) [237]. If policies are developed to promote the required uptake of 60% electric vehicle penetration by 2030 in order to meet this reduction, then electrical energy demand, through vechile battery charging, will increase. The CCC have also stated that *"there is scope for reducing industry emissions by almost half between now and 2030"* (relative to

2007 levels) [237]. In addition to energy efficiency improvements in buildings, this requires widespread adoption of alternative methods of heating such as heat pumps, thus also having the potential to increase electrical energy demand.

These observations motivate the following extension to this project: the case where future electricity demand increases substantially from current levels. In modelling terms, this would involve developing new daily and annual load profiles (e.g., Fig. 2.11), which are then used to calculate the full and residual annual load probability distributions. The methodologies outlined in Chapters 6 and 7 would then be applied in order to estimate the mix and amount of generation investment over time in response to policies promoting wind *and* high heat and transport electrification. Further, alternative load profile scenarios could be developed based on alternate usage patterns. For instance, if heat pumping and vehicle charging were to occur during peak periods (i.e, 07:00-19:00), then higher levels of installed capacity will be required relative to the base case. If demand growth is participated by the investor, then *ceteris paribus*, the simulated response of thermal generation investment may provide a similar level of security of supply risk to the base case. However the model could be used to establish the additional capacity requirements and increases in overall system costs, which are then fed back into the policy development process.

9.3.5 Other considerations

In this Section, some other possible areas of further work are summarised:

- Calculating appropriate levels of de-rated capacity margins for a given investment scenario: this model has used de-rated margin as a proxy for the level of generation adequacy risk. It should be noted however that the same levels of de-rated margin do not necessarily provide the same level of risk. It very much depends on the capacity mix and demand scenario being analysed. Further work looking at the uncertainty of the input parameters used to calculate expected de-rated margins and the level of underlying risk could be explored.

- Application to other power markets: this model has been applied to a GB market case study where high penetrations of wind power are expected to emerge. The model could be adapted and applied to other power systems where policies to promote different renewable sources of energy are being developed. Output from other technologies such as marine renewables could also be included when calculating residual load duration curves.

- Interactions between power markets: the dynamic investment model has been applied to a single electricity market. Assessment of the impact of renewable and generation adequacy policy in neighbouring regions is an active area of research (e.g., [128]). This is of particular interest across Europe where a combined European-wide network of transmission system operators (the ENTSO) has replaced the 6 existing European transmission system operators, thus signalling the imminent arrival of a single European power market.

9.4 Thesis conclusion

This thesis has taken an interdisciplinary approach to modelling generation investment in liberalised electricity markets. Many disciplines have called upon during this research. These include 1) dynamical systems and control theory to model the interactions between supply and demand; 2) rational expectations, Value at Risk and Prospect Theory when modelling investment decision processes; 3) statistical analyses when calculating generation adequacy risk; and 4) wind resource assessment when estimating GB onshore and offshore wind production.

This project has enhanced the body of knowledge surrounding investment cycles in the power sector, in particular where high penetration of wind power are likely to emerge. This thesis describes an implementation of a dynamic capacity market investment model, which includes an adapted method to calculate expected outputs, costs and revenues from thermal generation investments in a nonequilibrium market setting. Price mark-ups in oligopoly markets are accounted for. Also described is a method of using de-rated capacity margins as a proxy for security of supply risk in a power system with high penetrations of wind.

Like any simulation model, the results produced must be interpreted with a degree of caution. A number of assumptions and simplifications are required in order to create a tractable model, although many insights, such as those presented here, can still be gained. Going forward, this thesis can be used for knowledge transfer, and as a basis for exploring the topics of further research revealed as part of this research process.

References

[1] International Energy Agency, "Energy Market Experience; Lessons Learned from Liberalised Electricity Markets," tech. rep., 2005.

[2] D. Newbery, *Electricity Market Reform: An International Perspective*, ch. 3, pp. 109–144. Elsevier Science, 2006.

[3] S. Stoft, *Power System Economics Designing Markets for Electricity*. Wiley-Academy, 2002.

[4] P. B. D. Aikman, S. Kapadia, M. King, J. Proudman, T. Taylor, I. de Weymarn, and T. Yates, "Uncertainty in Macroeconomic Policy Making: Art or Science?," in *Paper delivered to The Royal Society Conference on "Handling Uncertainty in Science"*, Mar. 2010.

[5] D. S. Kirschen and G. Strbac, *Fundamentals of Power Systems Economics*. John Wiley & Sons, 2004.

[6] *Electricity Market Reform: An International Perspective*. Elsevier Science, 2006.

[7] S. Hesmondhalgh, "Is NETA the Blueprint for Wholesale Electricity Trading Arrangements of the Future?," *IEEE Trans. Power Systems*, vol. 18, pp. 548 – 554, May 2003.

[8] "GB Wholesale Electricity Market Liquidity: Summer 2010 Assessment," tech. rep., Ofgem, Jul. 2010. www.ofgem.gov.uk.

[9] P. L. Joskow, "Market Imperfections Versus Regulatory Imperfections," *CESifo DICE Report*, vol. 8, no. 3, pp. 3–7, 2010.

[10] P. D. Klemperer and M. A. Meyer, "Supply Function Equilibria in Oligopoly under Uncertainty," *Econometrica*, vol. 57, no. 6, pp. 1243–1277, 1989.

[11] R. Baldick, S. Kolos, and S. Tompaidis, "Interruptible Electricity Contracts from an Electricity Retailer's Point of View: Valuation and Optimal Interruption," *Operations Research*, vol. 54, pp. 627–642, Jul. 2006.

[12] R. Green, *Competitive Electricity Markets and Sustainability*, ch. "Investments in generation capacity". Cheltenham, 2006.

[13] P. Twomey, R. J. Green, K. Neuhoff, and D. M. Newbery, "A Review of the Monitoring of Market Power: The Possible Roles of TSOs in Monitoring for Market Power Issues in Congested Transmission Systems," working papers, Massachusetts Institute of Technology, Center for Energy and Environmental Policy Research, 2005.

[14] D. Hurlbut, K. Rogas, and S. Oren, "Protecting the Market from "Hockey Stick" Pricing: How the Public Utility Commission of Texas is Dealing with Potential Price Gouging," *The Electricity Journal*, vol. 17, no. 3, pp. 26 – 33, 2004.

[15] P. Holmberg and D. Newbery, "The Supply Function Equilibrium and its Policy Implications for Wholesale Electricity Auctions," *Utilities Policy*, vol. 18, no. 4, pp. 209 – 226, 2010. Designing Electricity Auctions.

[16] C. Batlle and P. Rodilla, "A Critical Assessment of the Different Approaches Aimed to Secure Electricity Generation Supply," *Energy Policy*, vol. 38, no. 11, pp. 7169 – 7179, 2010.

[17] R. Green and N. Vasilakos, "Market Behaviour with Large Amounts of Intermittent Generation," *Energy Policy*, vol. 38, no. 7, pp. 3211 – 3220, 2010. Large-scale wind power in electricity markets with Regular Papers.

[18] P. Cramton and S. Stoft, "Forward Reliability Markets: Less Risk, Less Market Power, More Efficiency," *Utilities Policy*, vol. 16, no. 3, pp. 194 – 201, 2008.

[19] National Grid, "Operational Data." www.nationalgrid.com/uk/Electricity/Data/Demand+Data/.

[20] S. Arango and E. Larsen, "Cycles in Deregulated Electricity Markets: Empirical Evidence From Two Decades," *Energy Policy*, vol. 39, no. 5, pp. 2457 – 2466, 2011.

[21] "UK Electricity Generation Costs Update," tech. rep., Mott MacDonald, Jun. 2010. www.decc.gov.uk/en/content/cms/about/ec_social_res/analytic_projs/gen_costs/gen_costs.aspx.

[22] M. Ezekiel, "The Cobweb Theorem," *The Quarterly Journal of Economics*, vol. 52, no. 2, pp. 255–280, 1938.

[23] UKERC, "Investment in Electricity Generation - the Role of Costs, Incentives and Risks," tech. rep., May 2007.

[24] International Energy Agency, "Projected Costs of Generating Electricity," tech. rep., 2010 Edition.

[25] Power Engineering International, "CCGT: Breaking the 60 per cent Efficiency Barrier." www.powerengineeringint.com/articles/print/volume-18/issue-3/features/ccgt-breaking-the-60-per-cent-efficiency-barrier.html.

[26] UK MARKAL 2007 Electricity Generating Technologies, ukedc.rl.ac.uk/data.html.

[27] National Grid, "Seven Year Statements," tech. rep. www.nationalgrid.com/uk/Electricity/SYS/.

[28] Department of Trade and Industry, "Future of Nuclear Power," tech. rep., May 2007.

[29] World Nuclear Association, "The Economics of Nuclear Power," 2010. www.world-nuclear.org/info/inf02.html.

[30] Nuclear Decommissioning Authority, "Our sites." www.nda.gov.uk/sites/.

[31] Greenpeace and GWEC, "Global Wind Energy Outlook 2010," tech. rep., Global Wind Energy Council, Oct. 2010.

[32] A. K. Dixit and R. S. Pindyck, *Investment under uncertainty*. Princeton University Press, 1994.

[33] R. Gross, W. Blyth, and P. Heptonstall, "Risks, Revenues and Investment in Electricity Generation: Why Policy Needs to Look Beyond Costs," *Energy Economics*, vol. 32, no. 4, pp. 796 – 804, 2010.

[34] E. R. Larsen and D. W. Bunn, "Deregulation in Electricity: Understanding Strategic and Regulatory Risk," *The Journal of the Operational Research Society*, vol. 50, no. 4, pp. pp. 337–344, 1999.

[35] "Matching Renewable Electricity Generation with Demand," tech. rep., The University of Edinburgh, Commissioned by the Scottish Executive, 2006.

[36] R. A. Brealey and S. C. Myers, *Principles of Corporate Finance*. McGraw-Hill, 1996.

[37] G. P. Harrison, *An Assessment of the Impact of Climate Change on Hydroelectric Power*. PhD thesis, School of Engineering, University of Edinburgh, 2000.

[38] K. G. Jechoutek and R. Lamech, *Private Power Financing - From Project Finance to Corporate Finance*. World Bank, Oct. 1995.

[39] E. S. Beckjord, *The Future of Nuclear Power - An Interdisciplinary MIT Study*. Massachusetts Institute of Technology MIT, 2003.

[40] Moody's, "Rating Symbols and Definitions," Oct. 2010. www.moodys.com/researchdocumentcontentpage.aspx?docid=PBC_79004.

[41] E.ON, "Annual Report and Accounts 2009." www.eon.com/en/downloads/EON_Financial_Report_2009_.pdf.

[42] Edf, "Finacial Report 2009." www.shareholders-and-investors.edf.com/results-and-publications/annual-reports-42724.html.

[43] RWE Npower, "Annual Report 2009." www.rwe.com/web/cms/en/110822/rwe/investor-relations/financial-reports/.

[44] Scottish & Southern Energy plc, "Annual Report 2010." www.sse.com/SSEInternet/uploadedFiles/Corporate_Responsibility/Our_Performance/Reports/Report_items/AnnualReport2010.pdf.

[45] Scottish Power, "Annual Report and Accounts 2006." www.scottishpower.com/Annualandinterimreports.htm.

[46] Centrica, "Annual Report and Accounts 2008." www.centrica.com/files/reports/2008ar/index.asp?pageid=86.

[47] S. S. Oren, "Generation Adequacy via Call Options Obligations: Safe Passage to the Promised Land," *The Electricity Journal*, vol. 18, no. 9, pp. 1040 – 6190, 2005.

[48] S. Stoft, "The Demand for Operating Reserves: Key to Price Spikes and Investment," *IEEE Trans. Power Systems*, vol. 18, pp. 470 – 477, May 2003.

[49] W. M. Smith, “The California Experience in 2000/2001: Drivers for Load Management in Deregulated Electricity Markets,” in *Market Design Conference 2001*, Jun. 2001.

[50] S. Blumsack, “How the Free Market Rocked the Grid,” *IEEE Spectrum*, pp. 42–47, Dec. 2010.

[51] “A Comparison of PJMs RPM with Alternative Energy and Capacity Market Designs,” tech. rep., The Brattle Group, Sep. 2009.

[52] J. Bowring, *Electricity Market Reform: An International Perspective*, ch. 4, pp. 451–479. Elsevier Science, 2006.

[53] M. L. Telson, “The Economics of Alternative Levels of Reliability for Electric Power Generation Systems,” *The Bell Journal of Economics*, vol. 6, no. 2, pp. 679–694, 1975.

[54] B. F. Hobbs, M.-C. Hu, J. G. Inon, S. E. Stoft, and M. P. Bhavaraju, “A Dynamic Analysis of a Demand Curve-Based Capacity Market Proposal: The PJM Reliability Pricing Model,” *IEEE Trans. Power Systems*, vol. 22, pp. 3 –14, Feb. 2007.

[55] M. P. Bhavaraju, B. F. Hobbs, and M.-C. Hu, “PJM Reliability Pricing Model - A Summary and Dynamic Analysis,” in *IEEE Power Engineering Society General Meeting*, pp. 1 –3, Jun. 2007.

[56] C. Vazquez, M. Rivier, and I. J. Perez-Arriaga, “A Market Approach to Long-Term Security of Supply,” *IEEE Trans. Power Systems*, vol. 17, pp. 349 –357, May 2002.

[57] M. Bidwell, “Reliability Options: A Market-Oriented Approach to Long-Term Adequacy,” *The Electricity Journal*, vol. 18, no. 5, pp. 11 – 25, 2005.

[58] L. M. Ausubel and P. Cramton, “Using Forward Markets to Improve Electricity Market Design,” *Utilities Policy*, vol. 18, no. 4, pp. 195 – 20, 2010. Designing Electricity Auctions.

[59] G. L. Doorman and A. Botterud, “Analysis of Generation Investment Under Different Market Designs,” *IEEE Trans. Power Systems*, vol. 23, pp. 859 –867, Aug. 2008.

[60] N. van Hulst, “Lessons Learned from Liberalised Electricity Markets,” Dec. 2004. RIETI Policy Symposium, Electricity Market Reform: Current State and Remaining Challenges, www.rieti.go.jp/en/events/04121501/report.html.

[61] P. E. O. Aguirre, C. J. Dent, G. P. Harrison, and J. W. Bialek, “Realistic Calculation of Wind Generation Capacity Credits,” in *Integration of Wide-Scale Renewable Resources Into the Power Delivery System, 2009 CIGRE/IEEE PES Joint Symposium*, pp. 1–8, Jul. 2009.

[62] F. P. Jenkin, “The Generating Standard and Planning Margin,” tech. rep., Central Electricity Generating Board, 1896.

[63] C. J. Dent, K. R. W. Bell, A. W. Richards, S. Zachary, D. E. G. P. Harrison, and J. W. Bialek, “The Role of Risk Modelling in the Great Britain Transmission Planning and Operational Standards,” in *IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*, pp. 325 –330, Jun. 2010.

[64] R. Billinton and R. N. Allan, *Reliability Evaluation of Power Systems. 2nd edition*. New York : Plenum Press, c1996.

[65] N. Screen, V. Parail, D. Sinclair, and O. Rix, “Electricity Market Reform Analysis of Policy Options,” tech. rep., Redpoint Energy, Dec. 2010. www.decc.gov.uk/en/content/cms/consultations/emr/emr.aspx.

[66] National Grid, “Winter Outlook Consultation 2010/11,” tech. rep. www.nationalgrid.com/uk/Gas/TYS/outlook/.

[67] Department of Energy and Climate Change, “Electricity Market Reform Consultation Document,” tech. rep., Dec. 2010. www.decc.gov.uk/en/content/cms/consultations/emr/emr.aspx.

[68] “Our Electricity Transmission Network: A Vision for 2020,” tech. rep., ENSG. www.ensg.gov.uk/index.php?article=126.

[69] B. Hasche, A. Keane, and M. O’Malley, “Capacity Value of Wind Power, Calculation, and Data Requirements: the Irish Power System Case,” *IEEE Trans. Power Systems*, vol. 26, pp. 420 –430, Feb. 2011.

[70] J. Vardi, J. Zahavi, and B. Avi-Itzhak, “Variable Load Pricing in the Face of Loss of Load Probability,” *Bell Journal of Economics*, vol. 8, no. 1, pp. 270–288, 1977.

[71] C. J. Dent and J. W. Bialek, “Non-Iterative Method for Modeling Systematic Data Errors in Power System Risk Assessment,” *IEEE Trans. Power Systems*, vol. 26, no. 1, pp. 120–127, 2010.

[72] A. Keane, M. M. C. J., Dent, B. Hasche, C. D. Annunzio, K. Dragoon, H. Holttinen, N. Samaan, L. Soder, and M. O. Malley, “Capacity Value of Wind Power,” *IEEE Trans. Power Systems*, vol. PP, no. 99, p. 1, 2010, in press.

[73] Personal communication C. J. Dent. Energy Institute, Durham University, UK.

[74] D. M. Newbery and M. G. Pollitt, “The Restructuring and Privatisation of Britain’s CEGB–Was It Worth It?,” *The Journal of Industrial Economics*, vol. 45, no. 3, pp. pp. 269–303, 1997.

[75] “Report on the Generation Security Standard,” tech. rep., Electricity Council, Sep. 1986.

[76] J. Evans and R. Green, “Why Did British Electricity Prices Fall after 1998?,” Discussion Papers 05-13, Department of Economics, University of Birmingham, Jul. 2005.

[77] D. Newbery, “Power Markets and Market Power,” *Energy Journal*, vol. 16, no. 3, pp. 41 – 66, 1995.

[78] R. J. Green and D. M. Newbery, “Competition in the British Electricity Spot Market,” *The Journal of Political Economy*, vol. 100, no. 5, pp. 929–953, 1992.

[79] J. Bower, “Why Did Electricity Prices Fall in England and Wales? Market Mechanism or Market Structure?,” Tech. Rep. 0401008, EconWPA, Jan. 2004.

[80] M. de Nooij, C. Koopmans, and C. Bijvoet, "The Value of Supply Security: The Costs of Power Interruptions: Economic Input for Damage Reduction and Investment in Networks," *Energy Economics*, vol. 29, no. 2, pp. 277 – 295, 2007.

[81] D. Newbery, "Implications of Current NETA Design," Apr. 2011. www.eprg.group.cam.ac.uk/implications-of-current-neta-design/.

[82] Personal communication J. Skea. UKERC, 58 Princes Gate, London, UK.

[83] L. R. Clarke, "New Electricity Trading Arrangements in England and Wales," in *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES*, vol. 2, pp. 1470 – 1472 vol.2, Oct. 2002.

[84] OFFER, "Review of Electricity Trading Arrangements: Proposals," Jul. 1998.

[85] Association of Electricity Producers. www.aepuk.com/about-electricity/electricity-market/.

[86] I. Marlee, "Open letter: Liquidity in the GB Power Market Update and Next Steps," Dec. 2010. www.ofgem.gov.uk/CustomPages/Pages/Publications.aspx.

[87] "Liquidity in the GB Wholesale Markets," tech. rep., Ofgem, Jul. 2010. www.ofgem.gov.uk.

[88] National Grid, "Short-term Operating Reserve General Description of the Service." www.nationalgrid.com/NR/rdonlyres/72D4386B-2027-474C-B281-2384F5B21A5E/40978/TR11STORGeneralDescriptionFinal.pdf.

[89] "Postnote no. 203: Security of Electricity Supplies," tech. rep., Parliamentary Office of Science and Engineering, 2003. www.parliament.uk/documents/post/pn203.pdf.

[90] Economic and Social Data Service, www.esds.ac.uk/international/access/dataset_overview.asp#desc_ieaei.

[91] J. W. Bialek and C. J. Dent and D. Eager, "Electricity Market Reform: Consultation Response on Behalf of Durham Energy Institute." Mar. 2011.

[92] National Grid plc, "Average Cold Spell (ACS) Correction." www.nationalgrid.com/uk/sys_08/default.asp?action=mnch2_15.htm&sNode=2&Exp=N.

[93] S. Zachary and C. J. Dent, "Probability Theory of Capacity Value of Additional Generation," Apr. 2011. www.supergennetworks.org.uk/filebyid/609/file.pdf.

[94] Scottish and Southern Hydro Electric, "Power From the Glens." www.sse.com/uploadedFiles/Content/11_About_Us/How_we_run_are_business/Community/PowerFromTheGlens.pdf.

[95] P. Grant and S. Skillings, "Decarbonising the GB Power Sector: Evaluating Investment Pathways, Generation Patterns and Emissions Through to 2030," tech. rep., Redpoint Energy, Sep. 2009. www.redpointenergy.co.uk/.

[96] "Project Discovery: Energy Market Scenarios," tech. rep., Ofgem, Oct. 2009. www.ofgem.gov.uk.

[97] “Powering the future,” tech. rep., Parsons Brinkerhoff Mapping our Low Carbon Path to 2050, Dec. 2009. www.pbpoweringthefuture.com.

[98] “Impact of Intermittency: How the Variability Could Change the Shape of the British and Irish Electricity Markets,” tech. rep., Pöyry, Jul. 2009. www.poyry.com/linked/group/study.

[99] “Impact of Intermittency: How the Variability Could Change the Shape of the British and Irish Electricity Markets. Description of methodology,” tech. rep., Pöyry, Sep. 2009.

[100] Committee on Climate Change, “Meeting Carbon Budgets - the Need for a Step Change Supporting Research,” tech. rep., Oct. 2009. www.theccc.org.uk/reports/1st-progress-report/supporting-research-.

[101] Environmental Change Institute, “Variability of UK Marine Resources,” tech. rep., Carbon Trust, Jul. 2005. www.carbontrust.co.uk.

[102] NETA, “Balancing Mechanism Reporting System.” www.bmreports.com/.

[103] K. Neuhoff, S. Skillings, O. Rix, D. Sinclair, N. Screen, and J. Tipping, “Implementation of EU 2020 Renewable Target in the UK Electricity Sector: Renewable Support Schemes,” tech. rep., Redpoint Energy, Jun. 2008. www.berr.gov.uk/files/file46778.pdf.

[104] D. Sinclair, R. Steele, J. Tipping, and N. Screen, “Dynamics of GB Electricity Generation Investment,” tech. rep., Redpoint Energy, May 2007. www.decc.gov.uk/en/content/cms/consultations/emr/emr.aspx.

[105] Personal communication Redpoint Energy, London, UK.

[106] M. Ventosa, Álvaro Baíllo, A. Ramos, and M. Rivier, “Electricity Market Modeling Trends,” *Energy Policy*, vol. 33, no. 7, pp. 897 – 91, 2005.

[107] J. A. Bloom and L. Gallant, “Modelling Dispatch Constraints in Production Cost Simulations Based on the Equivalent Load Method,” *IEEE Trans. Power Systems*, no. 2, pp. 598–611, 1994.

[108] G. B. Dantzig and P. Wolfe, “Decomposition Principle for Linear Programs,” *Operations Research*, vol. 8, no. 1, pp. 101–111, 1960.

[109] J. Nocedal and S. Wright, *Numerical Optimization*. Springer Series in Operations Research, 1999.

[110] P.-O. Pineau, H. Rasata, and G. Zaccour, “Impact of some Parameters on Investments in Oligopolistic Electricity Markets,” *European Journal of Operational Research*, vol. 213, no. 1, pp. 180 – 195, 2011.

[111] R. Green and N. Vasilakos, “The Long-Term Impact of Wind Power on Electricity Prices and Generating Capacity,” *CCP Working paper*, no. 11-4, 2010. www.uea.ac.uk/ccp/CCP+Working+Paper+11-4.

[112] D. W. Bunn and E. R. Larsen, “Sensitivity of Reserve Margin to Factors Influencing Investment Behaviour in the Electricity Market of England and Wales,” *Energy Policy*, vol. 20, no. 5, pp. 420 – 429, 1992.

[113] R. Green, “Did English Generators Play Cournot? : Capacity Withholding in the Electricity Pool,” Tech. Rep. 04-010WP, MIT Center for Energy and Environmental Policy Research, Apr. 2004.

[114] J. A. C. Acevedo and S. A. Aramburo, “Capacity Utilization in Electricity Market Behaviour: an Experimental Analysis,” in *27th International Conference of the System Dynamics Society, Albuquerque, New Mexico, USA*, Jul. 2009.

[115] A. Ford, *Modeling the Environment: an Introduction to System Dynamics Models of Environmental Systems*. Island Press, 1999.

[116] N. Ghaffarzadegan, “How Small System Dynamics Models can Help the Public Policy Process,” *System Dynamics Review*, pp. 1099–1727, Oct. 2010.

[117] A. Ford, “System Dynamics and the Electric Power Industry,” *System Dynamics Review*, vol. 13, pp. 1099–1727, 1997.

[118] D. W. Bunn and F. S. Oliveira, “Modeling the Impact of Market Interventions on the Strategic Evolution of Electricity Markets,” *Operations Research*, vol. 56, no. 5, pp. 1116–1130, 2008.

[119] T. Kadoya, T. Sasaki, S. Ihara, E. Larose, M. Sanford, A. K. Graham, C. A. Stephens, and C. K. Eubanks, “Utilizing System Dynamics Modeling to Examine Impact of Deregulation on Generation Capacity Growth,” *Proceedings of the IEEE*, vol. 93, pp. 2060 –2069, Nov. 2005.

[120] K.-O. Vogstad, *A System Dynamics Analysis of the Nordic Electricity Market: The Transition from Fossil Fuelled Toward a Renewable Supply within a Liberalised Electricity Market*. PhD thesis, Norwegian University of Science and Technology, Trondheim, 2004.

[121] F. Olsina, F. Garces, and H.-J. Haubrich, “Modeling Long-Term Dynamics of Electricity Markets,” *Energy Policy*, pp. 1411–1433, 2006.

[122] P. Visudhiphan, P. Skantze, and M. Ilic, “Dynamic Investment in Electricity Markets and Its Impact on System Reliability,” in *Market Design 2001 Conference, Stockholm*, pp. 91–110, Jun. 2001.

[123] T. Häni, J. W. Bialek, and R. Cherkaoui, “Diploma Project: Modelling Generation Capacity Margins in Market Environments as a Dynamic Control Problem,” tech. rep., University of Edinburgh / École Polytechnique Fédérale de Lausanne, Feb. 2005.

[124] A. Ford, “Cycles in Competitive Electricity Markets: a Simulation Study of the Western United States,” *Energy Policy*, vol. 27, no. 11, pp. 637–658, 1999.

[125] S. A. Arango, *Essays on Commodity Cycles Based on Expanded Cobweb Experiments of Electricity Markets*. PhD thesis, The System Dynamics Group, The University of Bergen, 2006. bora.uib.no/handle/1956/2027.

[126] S. He, L. Xiao, J. Zhang, G. Liu, and M. Zeng, “Capacity Mechanism Analysis Based on Dynamic Simulation of Generation Investment,” in *Service Operations and Logistics, and Informatics, 2008. IEEE/SOLI 2008. IEEE International Conference on*, vol. 1, pp. 265 –269, Oct. 2008.

[127] S. Gary and E. R. Larsen, “Improving Firm Performance in Out-of-Equilibrium, Deregulated Markets Using Feedback Simulation Models,” in *Energy Policy*, vol. 28, pp. 845–855, 2000.

[128] M. Cepeda and D. Finon, “Generation Capacity Adequacy in Interdependent Electricity Markets,” *Energy Policy*, vol. 39, no. 6, pp. 3128 – 3143, 2011.

[129] J. J. Sànchez, J. Barquìn, E. Centeno, and A. López-Pena, “A Multidisciplinary Approach to Model Long-Term Investments in Electricity Generation: Combining System Dynamics, Credit Risk Theory and Game Theory,” in *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pp. 1 –8, Jul. 2008.

[130] W. Lise and G. Kruseman, “Long-term Price and Environmental Effects in a Liberalised Electricity Market,” *Energy Economics*, vol. 30, no. 2, pp. 230 – 248, 2008.

[131] T. Jäger, S. Schmidt, and U. Karl, “A System Dynamics Model for the German Electricity Market: An Analysis of Economic and Environmental Policy Related Impacts on Electricity Prices and CO_2 Emissions,” in *International Conference on Policy Modeling, Ottawa, Canada*, Jun. 2009.

[132] A. Ford, “Simulation Scenarios for Rapid Reduction in Carbon Dioxide Emissions in the Western Electricity system,” *Energy Policy*, vol. 36, no. 1, pp. 443–455, 2008.

[133] A. Ford, K. Vogstad, and H. Flynn, “Simulating Price Patterns for Tradable Green Certificates to Promote Electricity Generation from Wind,” *Energy Policy*, vol. 35, no. 1, pp. 91–111, 2007.

[134] A. Lopez-Pena, E. Centeno, J. Barquìn, and J. J. Sànchez, “Capacity Mechanisms and Renewable Promotion in a Business Dynamic-Based Generation Expansion Model Applied to the Spanish System,” Dec. 2009. Presented at Workshop on “A Feedback View on the Electricity Sector: Behaviour, Policy models and Long-term Consequences”, HEC, University of Lausanne.

[135] A. Ford, “Waiting for the Boom: a Simulation Study of Power Plant Construction in California,” *Energy Policy*, vol. 29, no. 11, pp. 847 – 869, 2001.

[136] A. Cartea and M. G. Figueroa, “Pricing in Electricity Markets: a Mean Reverting Jump Diffusion Model with Seasonality,” Finance 0501011, EconWPA, Jan. 2005.

[137] A. Botterud, M. D. Ilic, and I. Wangensteen, “Optimal Investments in Power Generation Under Centralized and Decentralized Decision Making,” *IEEE Trans. Power Systems*, vol. 20, pp. 254 – 263, Feb. 2005.

[138] A. Botterud and M. Korpås, “A Stochastic Dynamic Model for Optimal Timing of Investments in New Generation Capacity in Restructured Power Systems,” *International Journal of Electrical Power & Energy Systems*, vol. 29, no. 2, pp. 163 – 174, 2007.

[139] P. J. H. Schoemaker and C. A. J. M. van der Heijden, “Integrating Scenarios into Strategic Planning at Royal Dutch/Shell,” *Strategy & Leadership*, vol. 20, no. 3, pp. 41–46, 1992.

[140] J. F. Muth, “Rational Expectations and the Theory of Price Movements,” *Econometrica*, vol. 29, pp. 315–335, Jul. 1961.

[141] S. Fischer, *Rational Expectations and Economic Policy*. The University of Chicago Press, 1980.

[142] E. Grunberg and F. Modigliani, “The Predictability of Social Events,” *The Journal of Political Economy*, vol. 62, pp. 465–478, Dec. 1954.

[143] The Economist, “The Economics A-Z.” www.economist.com/research/economics/alphabetic.cfm.

[144] K. J. Li, “Jensen’s Inequality,” *Mathematical Excalibur*, vol. 5, Sep. 2000.

[145] M. Friedman and L. J. Savage, “The Utility Analysis of Choices Involving Risk,” *The Journal of Political Economy*, vol. 56, no. 4, pp. 279–304, 1948.

[146] H. Jin and X. Zhou, “Greed, Leverage, and Potential Losses: A Prospect Theory Perspective.” people.maths.ox.ac.uk/~zhouxy/download/greed_mf.pdf.

[147] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, “Coherent Measures of Risk,” *Mathematical Finance*, vol. 9, no. 3, pp. 203 – 228, 1999.

[148] Y. Chang, “The New Electricity Market of Singapore: Regulatory Framework, Market Power and Competition,” *Energy Policy*, vol. 35, no. 1, pp. 403 – 412, 2007.

[149] M. H. Asgari and H. Monsef, “Market Power Analysis for the Iranian Electricity Market,” *Energy Policy*, vol. 38, no. 10, pp. 5582 – 5599, 2010.

[150] L. Granier and M. Podesta, “Bundling and Mergers in Energy Markets,” *Energy Economics*, vol. 32, no. 6, pp. 1316 – 1324, 2010.

[151] Office of Fair Trading, “Mergers: Substantive Assessment Guidance,” tech. rep., May 2003.

[152] “Structure and Performance of Six European Wholesale Electricity Markets in 2003, 2004 and 2005,” DG Comp, London Economics in association with Global Energy Decisions, Feb. 2007. ec.europa.eu/competition/sectors/energy/inquiry/electricity_final_part4.pdf.

[153] J. W. Bialek, “Gaming the Uniform-Price Spot Market: Quantitative Analysis,” *IEEE Trans. Power Systems*, vol. 17, pp. 768 – 773, Aug. 2002.

[154] F. E. Kydland and E. C. Prescott, “Time to Build and Aggregate Fluctuations,” *Econometrica*, vol. 50, pp. 1345–70, Nov. 1982.

[155] R. Cooper, J. Haltiwanger, and L. Power, “Machine Replacement and the Business Cycle: Lumps and Bumps,” NBER Working Papers 5260, National Bureau of Economic Research, Inc, Sep. 1995.

[156] D. Gale, “Delay and Cycles,” *Review of Economic Studies*, vol. 63, pp. 169–198, 1996.

[157] A. BarIlan, A. Sulem, and A. Zanello, “Time-to-Build and Capacity Choice,” *Journal of Economic Dynamics and Control*, vol. 26, pp. 69–98, Jan. 2002.

[158] A. BarIlan and W. C. Strange, “The Timing and Intensity of Investment,” *Journal of Macroeconomics*, vol. 21, pp. 57–77, Jan. 1999.

[159] M. Costeniuc, M. Schnetzer, and L. Taschini, “Entry and Exit Decision Problem with Implementation Delay,” *Journal of Applied Probability*, vol. 45, no. 4, pp. 1039–1059, 2008.

[160] P. Asea and P. J. Zak, “Time-to-Build and Cycles,” UCLA Economics Working Papers 767, UCLA Department of Economics, May 1997.

[161] M. Bambi, G. Fabbri, and F. Gozzi, “Optimal Policy and Consumption Smoothing Effects in the Time-to-Build AK Model,” Discussion Papers (IRES - Institut de Recherches Economiques et Sociales) 2010029, UniversitÈ catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES), Aug. 2010.

[162] R. Boucekkine, O. Licandro, and C. Paul, “Differential-Difference Equations in Economics: On the Numerical Solution of Vintage Capital Growth Models,” *Journal of Economic Dynamics and Control*, vol. 21, no. 2-3, pp. 347 – 362, 1997.

[163] I.-K. Cho and S. P. Meyn, “Efficiency and Marginal Cost Pricing in Dynamic Competitive Markets with Friction,” *Theoretical Economics*, vol. 5, pp. 215 –239, May 2010.

[164] F. L. Alvarado, “The Stability of Electric Energy Markets,” *Decision Support Systems*, vol. 24, no. 3-4, pp. 233–241, 1999.

[165] A. Syothert and I. MacLeod, “Competitive Bidding as a Control Problem,” *IEEE Trans. Power Systems*, vol. 15, pp. 88–94, Feb. 2000.

[166] G. Gutiérrez-Alcaraz and G. B. Sheblé, “Generation Companies Decision-Making Modeling by Linear Control Theory,” *Electric Power Systems Research*, vol. 80, no. 7, pp. 815 – 827, 2010.

[167] Y. F. Liu and F. F. Wu, “Risk management of Generators’ Strategic Bidding in Dynamic Oligopolistic Electricity Market Using Optimal Control,” *IET Generation, Transmission and Distribution*, vol. 1, pp. 388 –398, May 2007.

[168] W. Yang and G. B. Sheble, “Modeling Generation Company Decisions and Electric Market Dynamics as Discrete Systems,” in *Power Engineering Society Summer Meeting, 2002 IEEE*, vol. 3, pp. 1385 –1391 vol.3, Jul. 2002.

[169] M. Assili, D. B. Javidi, M. Hossein, and R. Ghazi, “An Improved Mechanism for Capacity Payment Based on System Dynamics Modeling for Investment Planning in Competitive Electricity Environment,” *Energy Policy*, vol. 36, no. 10, pp. 3703 – 3713, 2008.

[170] P. Giabardo, M. Zugno, P. Pinson, and H. Madsen, “Feedback, Competition and Stochasticity in a Day Ahead Electricity Market,” *Energy Economics*, vol. 32, no. 2, pp. 292 – 301, 2010.

[171] ELEXON Portal, “Market Data Dashboard (login required).” www.bsccentralservices.com/.

[172] Personal communication D. Newbery. EPRG, University of Cambridge, UK.

[173] OFGEM, “Security of Supply October 2003 to March 2004,” Sep. 2004. www.ofgem.gov.uk.

[174] R. Kannan, N. Strachan, S. Pye, G. Anandarajah, and N. Balta-Ozkan, “UK MARKAL Model Documentation,” tech. rep., 2007. www.ukerc.ac.uk.

[175] P. A. M. Dirac, *The Principles of Quantum Mechanics. 4th ed.* Clarendon Press, 1981.

[176] S. E. Niles and L. J. Seibolt, “The Complexities of Repowering,” *World Energy*, vol. 8, no. 4, pp. 42 – 44, 2005.

[177] R. Sioshansi and S. Oren, “How Good are Supply Function Equilibrium Models: an Empirical Analysis of the ERCOT Balancing Market,” *Journal of Regulatory Economics*, vol. 31, pp. 1–35, 2007.

[178] J. B. Bushnell, E. T. Mansur, and C. Saravia, “Vertical Arrangements, Market Structure, and Competition: An Analysis of Restructured US Electricity Markets ,” *American Economic Review*, vol. 98, no. 1, pp. 237–66, 2008.

[179] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation and Control*. Wiley, 1996.

[180] Digest of United Kingdom Energy Statistics (DUKES), “Digest of United Kingdom Energy Statistics 2009 Annex A: Energy and Commodity Balances, Conversion Factors and Calorific Values.” www.decc.gov.uk/en/content/cms/statistics/publications/dukes/dukes.aspx.

[181] Department of Energy and Climate Change, “First Annual Update of the Short Term Traded Carbon Values (June 2010).” www.decc.gov.uk/en/content/cms/what_we_do/lc_uk/valuation/valuation.aspx.

[182] Department of Energy and Climate Change, “Fuel Mix Disclosure Data Table.” www.decc.gov.uk/en/content/cms/statistics/energy_stats/fuel_mix/fuel_mix.aspx.

[183] R. J. Green, “Carbon Tax or Carbon Permits: The Impact on Generators’ Risks,” *Energy Journal*, vol. 29, pp. 67–89, 2008.

[184] National Grid, “Short-term Operating Reserve.” www.nationalgrid.com/uk/Electricity/Balancing/services/reserveservices/STOR/.

[185] E.ON UK, “Responding to Market Changes.” www.eon-uk.com/about/869.aspx.

[186] e-ROC online auctions, www.e-roc.co.uk/intro.htm.

[187] National Grid, “Winter Outlook Consultation 2009/10,” tech. rep. www.nationalgrid.com/uk/Gas/TYS/outlook/.

[188] British Energy, "Our Nuclear Power Stations." www.british-energy.com/pagetemplate.php?pid=82.

[189] RenewableUK, formerly the British Wind Energy Association. www.bwea.com/.

[190] Digest of United Kingdom Energy Statistics, "Chapter 5: Electricity, Tables 5.6 and 5.7." www.decc.gov.uk/en/content/cms/statistics/publications/dukes/dukes.aspx.

[191] APX-ENDEX, www.apxgroup.com.

[192] BERR, Quarterly Energy Prices, March 07 and Sept. 09, www.berr.gov.uk/energy/statistics/publications/prices/index.html.

[193] First Uranium, www.firsturanium.com.

[194] D. Eager, J. W. Bialek, and T. Johnson, "Validation of a Dynamic Control Model to Simulate Investment Cycles in Electricity Generating Capacity," in *IEEE Power and Energy Society General Meeting*, pp. 1–8, Jul. 2010.

[195] R. Moreno, G. Strbac, F. Porrua, S. Mocarquer, and B. Bezerra, "Making Room for the Boom," *Power and Energy Magazine, IEEE*, vol. 8, pp. 36 –46, Oct. 2010.

[196] UK Met office. Met office surface data users guide. badc.nerc.ac.uk/data/ukmo-midas/ukmo_guide.html.

[197] RenewableUK, formerly the British Wind Energy Association. www.bwea.com/offshore/index.html.

[198] Vestas Turbine Designs. www.vestas.com/en/wind-power-solutions/wind-turbines.aspx.

[199] J. G. McGowan, J. F. Manwell, and A. L. Rogers, *Wind Energy Explained*. John Wiley and Sons, 2002.

[200] Personal communication A. Sturt. Control and Power Group, Imperial College London, UK.

[201] S. Hawkins, *A High Resolution Reanalysis of Wind Speeds over the British Isles for Wind Energy Integration*. PhD thesis, Institute for Energy Systems, University of Edinburgh, 2011.

[202] S. Hawkins and G. P. Harrison, "A High Resolution Reanalysis for Wind Energy Research," *Wind Energy*, 2011 (in press).

[203] Met Office Unified Model. www.metoffice.gov.uk/science/creating/daysahead/nwp/um.html.

[204] S. Hawkins, D. Eager, and G. P. Harrison, "Characterising the Reliability of Production from Future British Offshore Wind Fleets," in *IET Renewable Power Generation Conference*, Sep. 2011.

[205] Y. Feng, P. J. Tavner, H. Long, and J. W. Bialek, "Review of Early Operation of UK Round 1 Offshore Wind Farms," in *IEEE Power and Energy Society General Meeting*, pp. 1 –8, Jul. 2010.

[206] OFGEM, "Ofgem Renewables and CHP Register (public reports)." www.renewablesandchp.ofgem.gov.uk/.

[207] H. Balériaux, E. Jamoulle, and F. Linard de Guertechin, "Simulation de l'Explotation d'un Parc de Machines Thermiques de Production d'Électricité Couplé à des Stations de Pompage," *Revue E de la Societé Royale Belge des Electriciens*, vol. 7, pp. 225–245, 1967.

[208] R. R. Booth, "Power System Simulation Model Based on Probability Analysis," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-91, pp. 62–69, Jan. 1972.

[209] G. Gross, N. V. Garapic, and B. McNutt, "The Mixture of Normals Approximation Technique for Equivalent Load Duration Curves," *IEEE Trans. Power Systems*, vol. 3, pp. 368–374, May 1988.

[210] J. G. Inon, *Long-run Probabilistic Production Costing-Based Analysis of Electricity Market Designs for Long-Run Generation Investment Incentives under Spot-Pricing and Transmission Open-Access*. PhD thesis, John Hopkins University, 2012, in preparation.

[211] B. F. Hobbs, J. G. Inon, and S. E. Stoft, "Installed Capacity Requirements and Price Caps: Oil on the Water, or Fuel on the Fire?," *The Electricity Journal*, vol. 14, no. 6, pp. 23 – 34, 2001.

[212] B. F. Hobbs and Y. Ji, "A Bounding Approach to Multiarea Probabilistic Production Costing," *IEEE Trans. Power Systems*, vol. 10, pp. 853 –859, May 1995.

[213] Department of Energy and Climate Change, "Communication on DECC Fossil Fuel Price Assumptions." www.decc.gov.uk/en/content/cms/about/ec_social_res/analytic_projs/ff_prices/ff_prices.aspx.

[214] ICE Futures Europe, "Report Center; End of Day Reports (UK Natural Gas & Newcastle Coal Futures)." www.theice.com/marketdata/reports/ReportCenter.shtml.

[215] The Guardian. "Budget 2011: Carbon Tax Brings Higher Electricity Bills and Nuclear Windfalls" www.guardian.co.uk/uk/2011/mar/24/carbon-tax-electricity-bills-nuclear-windfall.

[216] Department of Energy and Climate Change, "Energy and Emissions Projections." www.decc.gov.uk/en/content/cms/about/ec_social_res/analytic_projs/en_emis_projs/en_emis_projs.aspx.

[217] "Operating the Electricity Transmission Networks in 2020," tech. rep., National Grid, Jun. 2011. www.nationalgrid.com/uk/Electricity/Operating+in+2020/.

[218] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*. McGraw-Hill, 2001.

[219] A. Breipohl, F. N. Lee, J. Huang, and Q. Feng, "Sample Size Reduction in Stochastic Production Simulation," *IEEE Trans. Power Systems*, vol. 5, pp. 984 –992, Aug. 1990.

[220] B. Willems and J. Morbee, "Market Completeness: How Options Affect Hedging and Investments in the Electricity Sector," *Energy Economics*, vol. 32, no. 4, pp. 786 – 795, 2010.

[221] J. Endrenyi, *Reliability Modeling in Electric Power Systems*. Wiley, 1978.

[222] M. Evans, N. Hastings, and B. Peacock, *Another Look at Plotting Positions*. Wiley, 2000.

[223] OFGEM, "Transmission Access Review." www.ofgem.gov.uk/Networks/Trans/ElecTransPolicy/tar/Pages/Traccrw.aspx.

[224] J. Valenzuela and M. Mazumdar, "Cournot Prices Considering Generator Availability and Demand Uncertainty," *IEEE Trans. Power Systems*, vol. 22, pp. 116 –125, Feb. 2007.

[225] MathWorks, "System Identification Toolbox - Product Description." www.mathworks.co.uk/products/sysid/description1.html.

[226] Department of Energy and Climate Change, "Planning Our Electric Future: a White Paper for Secure, Affordable and Low-carbon Electricity," tech. rep., Jul. 2011. www.decc.gov.uk/en/content/cms/legislation/white_papers/emr_wp_2011/emr_wp_2011.aspx.

[227] "PJM Manual 18: PJM Capacity Market," tech. rep., PJM Capacity Market Operations, May 2011.

[228] PJM Interconnection, "2012/2013 RPM Base Residual Auction Results." www.pjm.com/~/media/markets-ops/rpm/rpm-auction-info/2012-13-base-residual-auction-report-document-pdf.ashx.

[229] Department of Energy and Climate Change, "DECC lays foundations for smart meters rollout." www.decc.gov.uk/en/content/cms/news/pn11_032/pn11_032.aspx.

[230] Sinclair Knight Merz, "Energy Storage and Management Study," tech. rep., Oct. 2010.

[231] Renewable Energy Foundation, "REF calls for Transparency in the UK Electricity Market." www.ref.org.uk/press-releases/177-data-transparency-in-the-uk-electricity-sector-2010-2011.

[232] Digest of United Kingdom Energy Statistics, "Chapter 5: Electricity, Table 5.10 (historic)." www.decc.gov.uk/en/content/cms/statistics/publications/dukes/dukes.aspx.

[233] Department of Trade and Industry, "Meeting the Energy Challenge A White Paper on Energy," tech. rep., May 2007. www.decc.gov.uk/en/content/cms/legislation/white_papers/white_paper_07/white_paper_07.aspx.

[234] Third package for Electricity & Gas Markets. ec.europa.eu/energy/gas_electricity/legislation/third_legislative_package_en.htm.

[235] L. R. Noyes, "Two-Area Probabilistic Production Costing by the Method of Bi-Variant Cumulants," *IEEE Trans. Power Apparatus and Systems*, vol. 102, pp. 433–443, Feb. 1983.

[236] D. Burke, *Accommodating Wind Energy Characteristics in Power Transmission Planning Applications*. PhD thesis, University College Dublin, 2010.

[237] Committee on Climate Change, "The Fourth Carbon Budget - Reducing Emissions Through the 2020s," tech. rep., Dec. 2010. www.theccc.org.uk/reports/fourth-carbon-budget.

Appendix A

Additional material

A.1 AIMMS source code for calculating COPT [73]

```
Prob(b) := 0;
Prob(0) := 1;

for (s in Stations) do
  ProbDown(b) := 0;
  ProbUp(b) := 0;

  for (b|Prob(b)) do
    ProbUp(b+CapSteps(s)) += Prob(b)*   AvailProb(StatType(s));
    ProbDown(b          ) += Prob(b)*(1-AvailProb(StatType(s)));
  endfor;
  Prob(b) := ProbUp(b)+ProbDown(b);
endfor;

CumProb(0) := Prob(0);
for (b in {1..Nsteps}) do
  CumProb(b) := Prob(b) + CumProb(b-1);
endfor;
```

A.2 Marginal cost equations

Using the conversion factors listed given in 5.2.1.1 and technology emission assumptions given in 5.2.1.2, the SRMC equations are given by:

$$
\begin{aligned}
SRMC_{gas} &= \tfrac{1}{\eta}(3.6 \cdot 9.48F_g + 185F_{car}) + v, \\
SRMC_{ccgt} &= 64.38F_g + 349F_{car} + 2.0, \\
SRMC_{ocgt} &= 87.49F_g + 474F_{car} + 2.7, \\
\\
SRMC_{coal} &= \tfrac{1}{\eta}(\tfrac{3.6F_c}{0.02512} + 330F_{car}) + v, \\
&= 409F_c + 942F_{car} + 1.0, \\
\\
SRMC_{nuc} &= \tfrac{8.9F_u}{24 \cdot 45\eta} + v + e, \\
&= 0.023F_u + 2.68
\end{aligned} \tag{A.1}
$$

where F_g is the gas price (£/therm), F_c is the coal price (£/kg), F_u is the uranium price (£/kg), V is in (£/MWh), F_{car} is the carbon price (£/kg), e is the cost to convert, enrich and fuel fabricate (CEFF) uranium which is assumed to be 2.5 £/MWh, 8.9kg of Uranium (U_3O_8) is needed to produce 1kg of CEFF uranium (U) and 45 MWd/kg is the assumed burn-up [29]. In this example, the thermal efficiencies, η, and variable O&M costs, V, are taken from Table 5.6.

To illustrate, given the fuel prices: gas: 60 p/therm (= 0.6 £/therm), coal: 70 £/T (= 0.07 £/kg), uranium 70 £/lb (= 155 £/kg) and carbon price 20 £/T (= 0.02 £/kg). The estimated SRMC are:

$$
\begin{aligned}
SRMC_{ccgt} &= \tfrac{1}{0.53}(3.6 \cdot 9.48 \cdot (0.6) + 185 \cdot (0.02)) + 2.0 &&= 47.6£/MWh, \\
SRMC_{ocgt} &= \tfrac{1}{0.39}(3.6 \cdot 9.48 \cdot (0.6) + 185 \cdot (0.02)) + 2.7 &&= 64.7£/MWh, \\
\\
SRMC_{coal} &= \tfrac{1}{0.35}(\tfrac{3.6 \cdot (0.07)}{0.02512} + 330 \cdot (0.02)) + 1.0 &&= 48.5£/MWh, \\
\\
SRMC_{nuc} &= \tfrac{8.9 \cdot (155)}{24 \cdot 45 \cdot 0.36} + 0.18 + 2.5 &&= 6.2£/MWh,
\end{aligned} \tag{A.2}
$$

A.3 Capital expenditure schedule

For simplicity, the expenditure schedule applied in (5.18) for the model results presented in Section 7.5 uses a lagged formulation with

$$M_i^x = zM_{i+1}^x \tag{A.3}$$

	-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nuclear	0.1	0.1	0.2	0.2	0.2	0.2	0.2
Coal			0.2	0.2	0.2	0.2	0.2
Wind					0.3	0.3	0.4
CCGT					0.2	0.4	0.4
OCGT						0.5	0.5

Table A.1: ***Capital expenditure schedule applied in (5.18) for the preliminary model results presented in Section 5.5.***

where $z = 0.8$ (i.e. the capital outlay increases by 25% each year).

A.4 DECC fuel assumptions

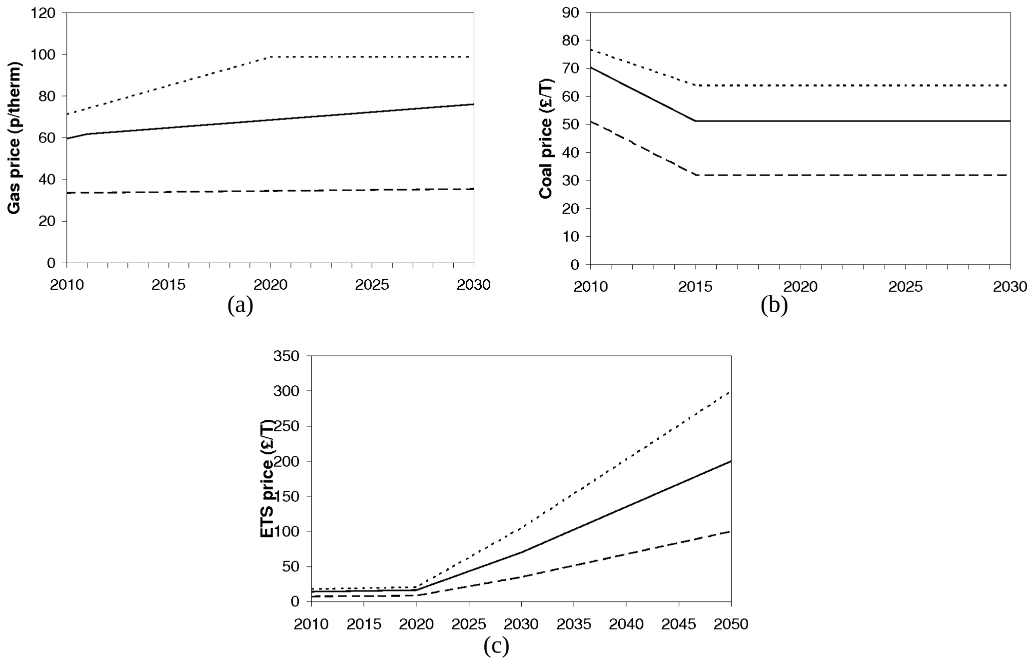


Figure A.1: *DECC low, central and high estimates for (a) gas, (b) coal and (c) ETS prices out to 2030 (ETS out to 2050). Sources: [181, 213]. All assumptions are in 2008 prices.*

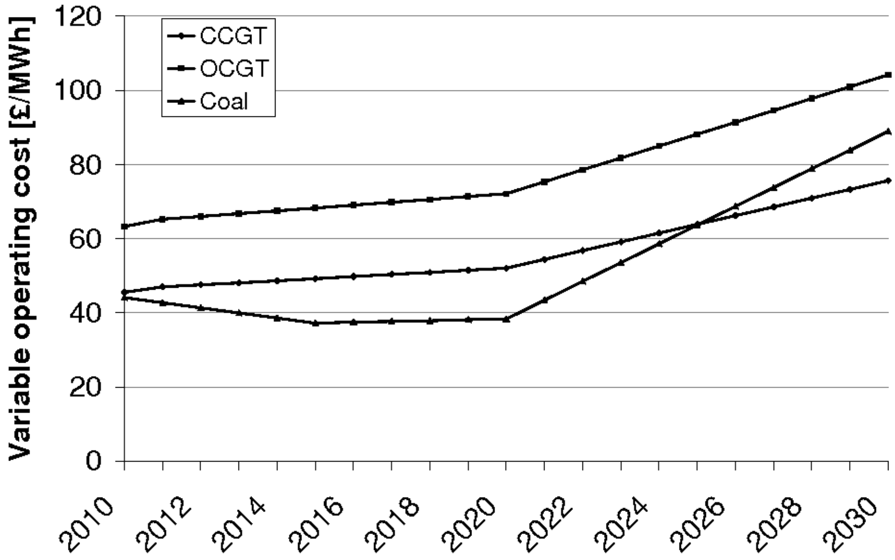


Figure A.2: *Plot of evolution of technology variable operating costs 2010-30 for central fuel cost assumptions shown in Fig. A.1. All values are in 2010 prices.*

A.5 Regional simulated monthly offshore wind generation

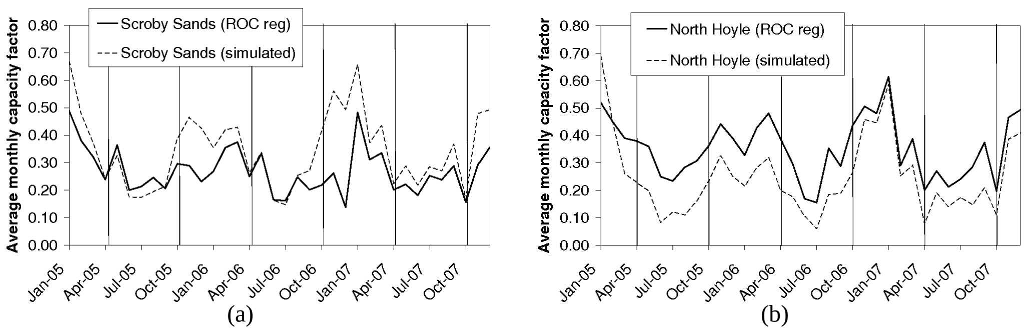


Figure A.3: *Plot of simulated monthly capacity factors versus ROC register data for (a) North Hoyle and (b) Scroby Sands offshore sites.*

A.6 Regional simulated monthly onshore wind generation

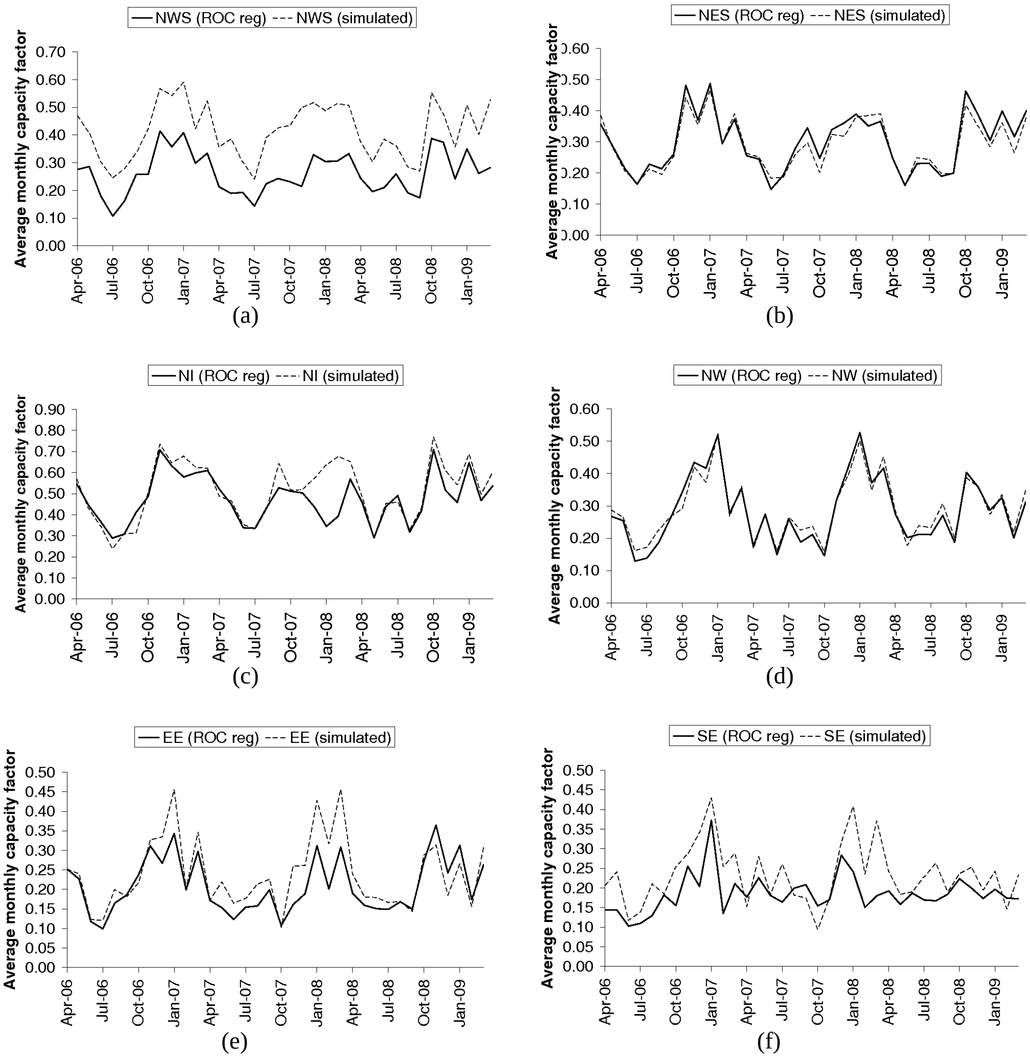


Figure A.4: *Plot of simulated capacity factors versus ROC register data for (a) North West Scotland, (b) North East Scotland, (c) North Islands, (d) North Wales, (e) East of England and (f) South of England.*

A.7 Proof of correlation property (7.20)

This can be proved by considering the correlation between $X + Y$ and X, where X and Y are independent variables, then

$$
\begin{aligned}
corr(X+Y,X) &= \frac{cov(X+Y,Y)}{\sqrt{Var(X+Y)\cdot Var(X)}} \\
&= \frac{E(([X+Y]-[E(X)+E(Y)])\cdot(X-E(X)))}{sqrtVar(X+Y)\cdot Var(X)} \\
&= \frac{E(([X-E(X)]+[Y-E(Y)])\cdot(X-E(X)))}{\sqrt{Var(X+Y)\cdot Var(X)}} \\
&= \frac{E([X-E(X)]\cdot(X-E(X))+[Y-E(Y)]\cdot(X-E(X))}{\sqrt{Var(X+Y)\cdot Var(X)}} \\
&= \frac{Var(X)+0}{\sqrt{Var(X+Y)\cdot Var(X)}} \\
&= \sqrt{\frac{Var(X)}{Var(X+Y)}}.
\end{aligned}
$$

Appendix B
Publications

The work described in this thesis has been reported in the following publications :

[B1] D. Eager, J. Bialek, and T. Johnson, 'Validation of a Dynamic Control Model to Simulate Investment Cycles in Electricity Generating Capacity', Proceedings of the IEEE Power and Energy Society General Meeting, Minneapolis, Minnesota, US, July 2010, pp. 391-394.

[B2] S. Hawkins, D. Eager and G. P. Harrison, 'Characterising the Reliability of Production from Future British Offshore Wind Fleets', Proceedings of the Renewable Power Generation Conference, IET, Edinburgh, UK, September 2011.

[B3] D. Eager, B. F. Hobbs, and J. W. Bialek, 'Dynamic Long-Term Modelling of Generation Capacity Investment and Capacity Margins: a GB Market Case Study', University of Cambridge, *Electricity Policy Research Group Working Paper Series*, no. 1201, January 2012.

[B4] D. Eager, B. F. Hobbs, and J. W. Bialek, 'Dynamic Modelling of Generation Capacity Investment in Markets with High Wind Penetration', *IEEE Trans. Power Systems*, March 2012 (in press).

Dynamic Modelling of Thermal Generation Capacity Investment: Application to Markets with High Wind Penetration

Dan Eager, *Student Member, IEEE,* Benjamin F. Hobbs, *Fellow, IEEE,* and Janusz W. Bialek, *Fellow, IEEE*

***Abstract*—Modelling the dynamics of merchant generation investment in market environments can inform the making of policies whose goals are to promote investment in renewable generation whilst maintaining security of supply. Such models need to calculate expected output, costs and revenue of thermal generation subject to varying load and random generator outages in a power system with high penetrations of wind.**

This paper presents a dynamic investment simulation model where the short-term energy market is simulated using probabilistic production costing using the Mix of Normals distribution technique to represent residual load (load net of wind output). Price mark-ups due to market power are accounted for. An 'energy-only' market setting is used to estimate the economic profitability of investments and forecast the evolution of security of supply. Simulated results for a Great Britain (GB) market case study show a pattern of increased relative security of supply risk during the 2020s. In addition, many new investments can recover their fixed costs only during years in which more frequent supply shortages push energy prices higher.

***Index Terms*—Power generation economics, Mix of Normals distribution, Thermal power generation, Wind power generation.**

I. Introduction

GOVERNMENTS who preside over liberalised energy markets are interested in how the mix and amount of generation investment will respond over time to policies promoting high penetration of variable output renewable power such as wind. Significant challenges arise when making an economic assessment of the potential for generating investments in a mixed thermal-wind system. Methods are needed that provide a reasonably accurate approximation of the expected contribution of generating units to serving varying load and the revenues received by doing so, whilst accounting for plant outages and variable renewable sources.

This paper demonstrates a probabilistic production costing method that considers the annual load curve and convolves it with generator outages using the Mix of Normals distribution (MOND) approximation. This is integrated into a dynamic capacity investment simulation. This production costing method was first described in [1] and then extended and used for the calculation of equilibrium capacity investment in a power market in [2] and [3], respectively. In this study, the method is applied for the first time to a nonequilibrium setting as part of a dynamic market simulation.

The second focus of the work is to assess the impact of high penetrations of wind power on risks associated with investment in conventional thermal generation. Therefore the method above is extended to include a residual load calculation (load net of wind output) using empirical load and simulated wind data. This residual load data is then used in the MOND production costing model. Finally, the MOND model is incorporated in a dynamic investment model and applied to a simplified Great Britain (GB) power system test case for an assumed (exogenously increasing) installed wind capacity.

The goal of this research is to address concerns about whether 'energy-only' markets (i.e., without capacity mechanisms) with high penetrations of wind are capable of inducing timely generation investments over a long-term time frame. Examples of 'energy-only' markets currently operating include GB, Australia's National Electricity Market, Alberta, Nordpool, Ontario and ERCOT.[1] This is of particular interest to policy makers whose goals are to promote investment in renewable generation, maintain an adequate level of resources, and reduce customer costs.

Building on the preliminary work of [4], the dynamic model employs classical control theory to capture the interactions between electricity supply and demand. It shares similarities with existing dynamic models of merchant generation investments (e.g., [5], [6]), however this particular application to a market with a high wind penetration is unique. Furthermore, the production costing methods used by previous dynamic models are deterministic, and therefore underestimate average costs (due to Jensen's inequality [7]).

This paper is organised as follows: in Section II, the dynamic investment model in which the MOND technique is embedded is described. Features of the investment decision are provided in Section III. In Section IV properties of a MOND is described and its application to a market situation is given. Section V describes input assumptions and the wind models used. For the purposes of illustration, a number of simplifying assumptions are made that in an actual application should be subjected to careful validation. Results from the dynamic simulation of capacity investment in a case study system with an existing installed capacity similar to that of GB are given

D. Eager is with the Institute for Energy Systems, The University of Edinburgh, EH9 3JL, UK, d.eager@ed.ac.uk

B. F. Hobbs is with the Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA, bhobbs@jhu.edu.

J. W. Bialek is with the Energy Group, School of Engineering, The University of Durham, DH1 3LE, UK. Janusz.Bialek@durham.ac.uk

[1] Note that some regulated vertically integrated utilities in US power markets (e.g., Midwest ISO) may be compelled by regulators to satisfy capacity margin criteria. In these circumstances investment decisions are not primarily governed by energy market prices.

in Section VI, Section VII presents conclusions.

II. THE INVESTMENT MARKET MODEL: OVERVIEW

Techniques from control theory are used to model market investment dynamics (Fig. 1). Because the model is dynamic, current market conditions (e.g., capacity under construction or retirements), and market price predictions are fed back to the investment block, modifying the investment behaviour. The resulting investment decisions are then fed back to the market, hence closing the loop.

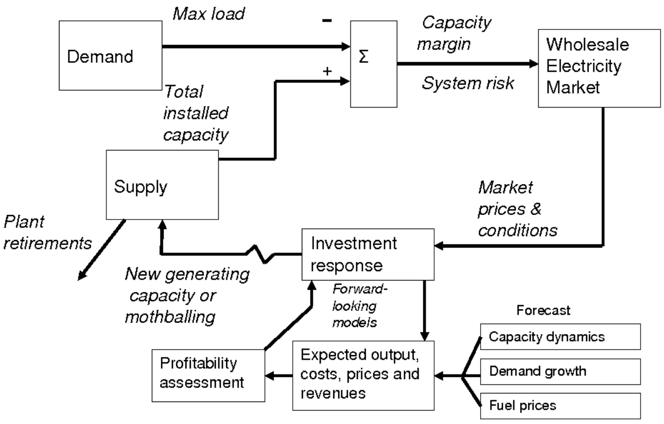


Fig. 1. Modelling generation capacity investment as a control problem; investment can be viewed as a negative feedback control mechanism with current and future energy prices (as a function of generation capacity margin) acting as a feedback signal.

The dynamics of the system concern the evolution of installed generation capacity and rate of demand growth. The rate of change in capacity at a particular time-step depends on new plant coming online or being de-mothballed, net of any plant being retired or mothballed [8]. Both are delayed signals from some earlier time. In the case of new plant, this delay is the lead time for construction, and for retiring plant, this delay is the design lifetime.[2] Mothballing requires zero delay, whilst de-mothballing requires a delay that is significantly less than full construction. An aggregate approach is taken whereby capacity is combined into five technology categories, namely nuclear, coal, wind, CCGT and OCGT, each with its own financial and technological characteristics. The model does not consider the transmission network and is conceptually a single bus system.

III. GENERATION CAPACITY INVESTMENT

A. General assumptions

In this model the investment decision is taken annually and is considered an irreversible decision. It is based on the Net Present Value (NPV) of anticipated future profits.

For simplicity, a single investor who is well acquainted with the structure of the market and capable of securing the necessary debt to finance large-scale plant investment is modelled. This representative agent approach has been used by other dynamic capacity market models (e.g., [9]). When estimating the profitability of an investment, a Monte Carlo (MC) approach is taken to obtain a probability distribution of profitability whereby estimates for conventional plant already under construction (including delays), demand growth, and fuel prices are considered stochastic.

In the base case, generators assume all plant will come online with 100% certainty, and remaining build time is stochastic. This is modelled as the sum of the expected build time plus a random variable (r.v) that is sampled from a lognormal distribution, $lnN(\mu, \sigma^2)$. Decision making under uncertainty is modelled by taking a risk averse attitude to investment (cf. Section III-B). Currently operational plant is expected to generate for the duration of its design lifetime.

Investors consider annual load growth to be stochastic and is sampled from a Normal distribution.

The present value of an investment in technology x (i.e., nuclear, coal, CCGT, OCGT) at any given decision year is given by:

$$V_x = \sum_{i=\tau_x}^{\alpha_x} \frac{GM_x^i - FC_x}{(1+r)^{i-\tau_x}} - (IC_x + DC_x) \tag{1}$$

where r is the firm's weighted average cost of capital (WACC), GM_x^i is the gross margin (cf. (11) or (16) depending on market bidding assumptions) for year i, FC_x is the generator fixed operating costs, and τ_x and α_x are the expected build time and plant lifetime, respectively. Although the investor randomly samples capacity construction times for plant already in the build stage, for simplicity we assume that investors consider only fixed build times when assessing the present value of an investment (1). In the case of plant lifetimes, the investor assumes that capacity will close at the end of its design life. A more sophisticated representation would have the investor consider the possibility of random construction times and lifetimes, which we leave for future research. Both these elements are modelled as fixed in the real-time simulation. We note that by assuming a fixed design lifetime, the model overlooks the option for generators to repower their existing plant. However, representing this additional choice for investors would add significant complexity to the model and increase computational burdens, and so is not considered here. IC_x and DC_x are the present worth of the investment cost and decommissioning cost, respectively. Only nuclear projects have considerable decommissioning costs; in the case of other plant types the decommissioning liabilities are assumed to be offset by the salvage value of the assets [10]. All cashflows are discounted to the start of the first year of operation.

Only the first n years of expected revenues are stochastically simulated by the investor (here $n = 7$); for the remaining years the (discounted) average of the simulated revenues are used (e.g., similar to [9]). Furthermore, investors cap the total expected annual revenues received from scarcity rents (see Section IV-B) at the annualised cost of an OCGT. These actions ensure that expectations about future revenues are not unduly influenced by high forward simulated wholesale prices due to generation retirements far in the future. No regulatory price caps are implemented in the real-time (annual)

[2] Further delays are caused by the need for investors to accumulate data in order to form their expectations about future market conditions.

simulation.

Because the model randomly samples capacity construction times, fuel prices and load growth, a large sample is required in order for investors to obtain reliable estimate of expected project value (1).

B. VaR criterion

Value at Risk (VaR) is a common criterion used in finance when investors place a high priority on avoiding poor outcomes [11]. Generally speaking, given a project value V, and defining the level of risk aversion by q, the VaR is defined as the value v^q such that $p(V \leq v^q) = 1 - q$. In this model, the distribution of V_x in (1) is computed by MC simulation of the stochastic variables. We model a risk averse investor with $q = 5\%$, and an investment is deemed attractive if $p(V_x > 0) \geq 95\%$.

The exact methods that generation companies actually use to evaluate investments under risk are uncertain, and whether they use probabilistic approaches is disputable. However, for the purposes of this simulation model, a probability model provides a convenient means for representing risks and how greater risk discourages investment.

Despite their popularity, VaR models do have a limitation relative to other risk criteria (such as Min Max Regret) as VaR analysis must consider probabilities of outcomes, but typically these probabilities are not well known. An alternative approach would be to use Min Max Regret whereby an investment decision for a particular realisation of the system is analysed. Regret is defined as the difference between the net benefit incurred when an investment decision is made for a realised state of nature, and the maximum net benefit across all feasible strategies that can be obtained under that realised state. For instance, investing in nuclear generation for a realised carbon price that is lower than expected might have a higher regret than investing in gas-fired generation under the same price. However, the Min Max Regret criterion only considers the worst regret for each alternative, which makes its results very sensitive to the exact set of scenarios that are considered.

C. Modelling aggregate investment response

In some circumstances the expected profitability of new investments is extremely high, thus triggering a wave of new builds. In such cases, the investment rate will be limited by: 1) the firm's beliefs about how many other market participants will move to invest, 2) on the impact of new investment on the profitability of their existing plant, and 3) on the ability of the firm to secure the debt required to fund multiple projects [6]. Using an aggregate investor response curve (AIRC) is useful in models of this type. For instance, in [9], the aggregate investment rate with increasing in the *"risk-adjusted forecast profit"*, which is derived from the investor's (risk averse, concave) utility function.

In this paper, a function is applied to the outcome of the VaR decision rule in order to estimate the aggregate investment response of the market. This function is increasing with the expected profitability and is given by:

$$\xi_i = max \left\{ 0, \xi_{max} \cdot \left(1 - e^{(-\beta \cdot PI_x^q)} \right) \right\}, \qquad (2)$$

where

$$PI_x^q = \frac{v_x^q}{(IC_x + DC_x)} \qquad (3)$$

is the Profitability Index (PI) and ξ_{max} is the maximum yearly investment lump per technology x. ξ_{max} and β are calibrated so zero investment is made if $PI_x^q < 0$, and $\xi_i = i_x$ volume of investment is made if $PI_x^q = 1$, where i_x is a fixed constant. Note that ξ_i determines the output of the investment response block in the feedback system depicted in Fig. 1.

The function used in the base case is shown in Fig. 2 with fixed $i_x = 2$ GW and $\xi_{max} = 4$ GW, and $\beta = 0.7$ resulting from the calibration. Changing β alters the aggregate response, as shown by the dotted lines. For multiple investments with $PI_x > 0$, the option with the highest PI_x is chosen.

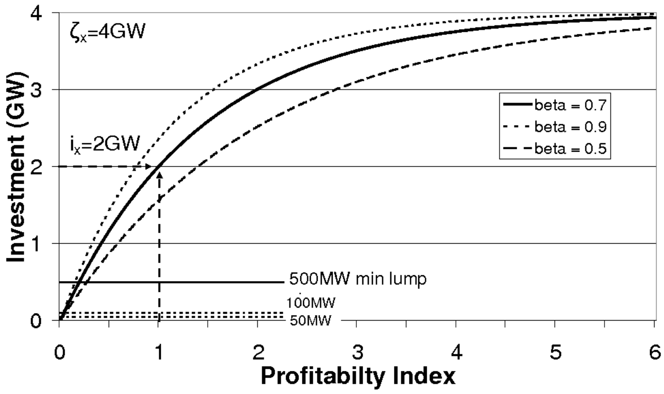


Fig. 2. Plot of model aggregate investment response curve defined by (2) (solid line) where $i_x = 2\ GW$ and $\xi_{max} = 4\ GW$. Also shown are the minimum investment lump sizes along with curves for different values of β.

The degree of heterogeneous investment response is not modelled directly, however the investment response of the model is a smooth function of prices and costs, reflecting how a heterogeneous group of market participants would likely respond to changes in market conditions. In particular, the amount of investment is a smooth monotonic function of expected returns, as reflected in the AIRC (2), i.e., it is not a type of response where if profits exceed a threshold, a large amount of investment occurs.

Furthermore, forward and bilateral contracting between generators and load serving entities is also not modelled explicitly, however both forward prices and our investors' expectations are driven by the same market factors. Moreover those contracting are implicit in the way investor price expectations are modelled. In a commodity market, forward price expectations will be driven by the same economics as short-term prices. For instance, in the face of capacity retirement and demand growth, forward prices will rise. Our modelled investor simulates prices 7 years out, which is further ahead than most forward electricity markets go. Moreover explicitly representing contracts and forward price expectations would make the model significantly more complicated. For instance, [12] presents a single stage investment problem in the presence of endogenous contract and physical markets. Even in that very simple circumstance, the model and its analysis is very complex. The effect of more long-term forward and bilateral contracts would be to reduce revenue risk for investors, i.e., lowering risk aversion, thus making them more willing to invest for a given distribution of energy prices. Therefore the effect of more or less forward contracting can be implicitly considered by adjusting the risk aversion parameter of the VaR model (i.e., the level of risk aversion, q, or the WACC, r, in (1)).

IV. Production costing by Mix of Normals

When estimating expected profits, varying loads (e.g., in the form of the load duration curve (LDC)), the expected contribution of generating units to serving these loads, and the revenues they receive by doing so must all be approximated. It is helpful if the technique used is computationally fast, accurate and robust. A mixture of Normals distribution (MOND) approximation of the LDC meets this requirement and is also straightforward to convolve with plant outages for the expected output calculations.

A MOND is described as follows: consider a set, $Y = \{y_1, \ldots, y_n\}$, of Normally distributed r.vs. with the i^{th} element having mean μ_i and variance σ_i. Let $\Phi(x|\mu_i, \sigma_i)$ be the cumulative distribution function (cdf) of y_i. A MOND is a convex combination of the Normal distributions and is defined by

$$F(x) = \sum_{i=1}^{n} p_i \Phi(x|\mu_i, \sigma_i), \quad (4)$$

with $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$, where p_i is the weight of the component y_i [1]. Note that if X and Y are two independent random variables each with MOND $F_X(x)$ and $F_Y(y)$ given by (4), then $X+Y$ is a MOND.[3] In this application the distribution of load (a MOND) is convolved with the distribution for available conventional thermal generation (cf. Section IV-A). It is a standard assumption in probabilistic costing and loss-of-load probability models that the outages of different generating units are independent of each other, and independent of load [13], [14].

A MOND fit for approximating the annual LDC (MW) is required. For example, if $f_L(x)$ is the probability density function of load and $F_L(x)$ is the cumulative distribution function of $f_L(x)$, then the LDC is simply the rotated and rescaled load *exceedence* distribution.[4] This is the inverse of $8760(1 - F_L(x))$ where

$$F_L(x) = \sum_{k=1}^{K} p_k \Phi_k(x|\mu_k, \sigma_k), \quad (5)$$

which is a mixture of K Normals (Φ_k) with the same properties as (4). For a particular K, the best fitting value of each μ_k, σ_k and p_k can be found by solving an optimization problem that minimizes the sum of squares of the difference between observed and fitted values of the LDC.

To illustrate the accuracy of this technique at approximating a LDC, Fig. 3 shows the distribution of the GB hourly loads for the period 2005-09 (normalised by the year's average demand) and the fitted distribution. A normalisation is necessary when comparing multiple years to account for demand growth. The difference between the LDC plots in Fig. 3(b) is not visible at this scale owing to the excellent fit provided by the MOND.

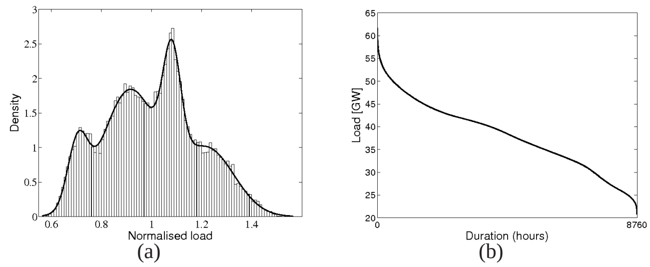


Fig. 3. (a) MOND pdf (4 Normal components clearly visible by their distinct peaks) and histogram of normalised load data and (b) LDC fit with negligible visible difference between mix of 4 Normals and empirical data.

A. MOND with conventional thermal generation (after [1]) and extension for wind

The available capacity at each hour from a particular unit is a r.v. which is characterised by the unit's Forced Outage Rate (FOR). Let the capacity of unit u be defined by c_u, its FOR by ρ_u and expected available capacity $G_u = c_u(1-\rho_u)$. That is, the distribution of the unit's available capacity follows a Bernoulli distribution between zero and c_u.[5]

If m_u units of type n share the same capacity and FOR characteristics and are subject to independent forced outages, they can be treated as a single pseudo-unit (or generator) with a distribution with the following moments:

$$E(G_n) \quad = \mu_n \quad = m_u c_u (1 - \rho_u), \quad (6)$$
$$Var(G_n) \quad = \sigma_n^2 \quad = m_u c_u^2 \rho_u (1 - \rho_u), \quad (7)$$

where (pseudo-) generator, G_n, has capacity $c_n = m_u c_u$. To simplify the presentation for the remainder of the paper, the convention will be to use 'generator' when referring to a 'pseudo-unit' (collection of units of a given type), and c_n when referring to the capacity of that generator, because this is the last time individual units will be discussed.

Now the convolution property of the MOND is used to determine the *effective load duration curve* (ELDC) facing the next generator to be dispatched (in merit order) [14]. If the units can be grouped into N (pseudo-) generators[6] each with characteristics (6) and (7), then we can define

$$L_n(x) = P\left\{L - \sum_{i=1}^{n} G_i > x\right\} \quad (8)$$

as the load minus the available capacity of generator types $1 \ldots n$ $(1 \leq n \leq N)$ [1]. If $F_n(x) = 1 - L_n(x)$ is the cumulative probability of effective load $x = L - \sum_{i=1}^{n} G_i$ facing the $(n+1)^{th}$ generator. $F_n(x)$ is computed by convolving the MONDs for load and available capacity.

[3] The proof of this property is given in appendix A of [1].

[4] The exceedence distribution gives $P(X > x)$, that is the probability that the r.v. X (in this case load) is greater than x.

[5] More general models exist to account for partial unit outages (e.g., see [13]), however they are not considered here.

[6] The more units in a group, the closer the Binomial distribution is to Normal.

The expected energy served e_n (in MWh/yr) by generator type n can then be given by

$$E[e_n] = 8760 \int_0^{\infty} \left[L_{n-1}(x) - L_n(x)\right] dx, \qquad (9)$$

where $L_{n-1}(x)$ is the load still to be met after adding generator type $n-1$ and $L_n(x)$ is the load still to be met after adding generator n, which at the start of the convolution process ($n = 0$) is obtained from (5).

The distribution for the ELDC after convolving in n generators is given by:

$$L_n(x) = 1 - \sum_{k=1}^{K} p_k \Phi_k(x|\mu_{L_k} - \sum_{i=1}^{n} \mu_{G_i}, \left[\sigma_{L_k}^2 + \sum_{i=1}^{n} \sigma_{G_i}^2\right]^{\frac{1}{2}}). \qquad (10)$$

Thus, the ELDC is described by a MOND with the same number of component Normals as the original LDC.

Given each $L_n(x)$ ($n = 1 \ldots N$), the probability that generator n or higher (in the merit order) is the marginal source of energy (and so sets the price) is given by $L_n(0)$, and $h_n = L_{n-1}(0) - L_n(0)$ is the probability that generator n is on the margin.

The annual expected energy unserved (EEU) can be calculated by integrating $L_{N+1}(x)$ (i.e., after convolving in the complete set of N generators) from 0 to ∞ and multiplying by 8760 hrs/yr. Furthermore the Loss-of-Load Expectation (LOLE) for the period is determined by computing $8760 \cdot L_{N+1}(0)$ [13].

Unlike conventional thermal generation units whose individual availabilities are assumed to be independent, wind generators rely on the availability of the "fuel" resource and therefore a dependency between available wind generation at different wind plants is introduced. To address this issue, an exogenous wind capacity is assumed and the resulting residual LDC facing thermal generation is computed and the MOND approximation is applied to this wind-adjusted data set. The residual load in each hour is simply the hourly load minus the sum of hourly output from all wind plants. By taking this approach, the resulting distribution of net load takes into account both spatial correlations and temporal (e.g., monthly and diurnal) trends in the availability of the wind resource and its relationship with demand.

Further the residual load approach allows us to calculate the number of hours that available wind generation is greater than aggregated demand, i.e., those hours when wind sets the system marginal price. This is computed using the residual load exceedence distribution before convolving any of the available thermal generation, i.e., the inverse of $8760(1 - F_L(0))$ where $F_L(x))$ is given by (5). This is of particular interest in systems with high penetrations of wind when direct wind dispatch may be required to curtail production at times when available wind exceeds demand.

Wind production may need to be curtailed under other circumstances. In particular, our implementation of the MOND technique does not consider the possibility of available wind generation exceeding either 1) available export capacity in a generation pocket due to transmission congestion or 2) raw demand net of inflexible base load (e.g., nuclear). Nor can the load duration curve method (of which the MOND is a particular case) consider the possibility that ramp rate limitations could also result in wind spill. Considering each in turn, given that the model is single bus, number 1) cannot be addressed here and is left for future research.[7] In this application, it is assumed that UK Government policy will ensure that the amount of congestion in the future will not be so large as to affect the basic economics of thermal generation investment. This is consistent with the GB regulator's 'connect and manage' transmission access policy [15]. In the case of number 2), if inflexible base load generation needs to be kept running then this will affect the economics of the wind generator being constrained (i.e., the wind generator is given a congestion payment by the System Operator). However, installed wind capacity is an exogenous model parameter in this analysis. The presence of inflexible generation can also affect the economics of baseload generation by increasing the number of hours of zero or negative prices. However, the amount of hours when this occurs is relatively small in most of our simulations. For instance, for the GB case study presented in Sections V-VI, under the assumption that all plants can be turned down except nuclear, in no year and in no scenario is the probability of the net demand being below the expected available nuclear capacity greater than 7%. Future applications of this method could approximate the effect of inflexible generation by dispatching its must run capacity first and assuming a zero or negative price for the portion of the time that this capacity is on the margin.

It is important to note that the residual LDC approach removes the chronology of the wind and load time series. This can be an important omission in the presence of large amounts of hydro and pumped hydro generation; in such cases, chronological production costing methods may be preferred to load duration curve methods. However our implementation of the MOND technique is applied to the GB power system where the amount of hydro and pumped hydro is relatively small (about 4% of capacity), so the use of a load duration curve approach is reasonable.[8]

B. Expected revenues from the energy market

During a particular year, the probability that generator n will be at the margin is given by h_n and the price in that event will be the marginal cost of the generator, MC_n. This assumes price-taking (competitive) behaviour by generators. Furthermore, once the convolution process has been completed for all N generators, the probability that there will be insufficient generation to meet demand is given by $L_{N+1}(0)$ and under this condition the price is assumed to reach Value of Loss Load (VOLL). The expected gross margin for a particular MW of capacity belonging to generator n when generator i is at the margin is given by: $R_n^i = max\{h_i(\pi_i - MC_n), 0\}$ (£/MWh), where π_i is the wholesale price when generator

[7] If a representation of the load and available wind production in each region of the network is available, then curtailment of wind production due to transmission congestion could in theory be assessed by multi-area production costing methods.

[8] Note that there is little scope for new build of hydro technologies in GB due to the lack of suitable sites.

i is at the margin, which in the absence of market power is given by MC_i. The expected annual perfectly competitive gross margin (revenue minus variable cost) used in (1) can be calculated by (£/MW/yr):

$$CGM_n = 8760(1-\rho_n)\Big[\sum_{i=1}^{N} R_n^i + (h_{N+1}(VOLL - MC_n))\Big]. \quad (11)$$

It is assumed that in a competitive energy market, generators bid to produce electricity at or around their marginal cost. But recent analyses of the GB market [16] showed a tendency for balancing market (BM) prices to rise above the estimated marginal cost of the last generator dispatched in peak demand hours. Empirical analyses from other markets support this claim [17], [18]. This price mark-up during peak demand hours occurs because firms can raise their bids knowing that the lack of alternative resources will mean their bids will be accepted. Accordingly, in this study price, π, includes an additional mark-up term that alters the shape of the aggregate supply curve as the system approaches scarcity. This new price function is described in [4] and is defined here as:

$$\pi(L, G_1, G_2, \ldots, G_N) = mc(L, G_1, G_2, \ldots, G_N) + w(L, G_N^*) \quad (12)$$

where $mc(L, G_1, G_2, \ldots, G_N)$ is the marginal production cost of meeting the load, L, given realised total available generation $G_N^* = \sum_{i=1}^{N} G_i$. The second term, $w(L, G_N^*) = ae^{b(L-G_N^*)}$, is a function of the capacity margin, defined as total available capacity minus load in a particular hour. The parameters a and b are calibrated so that a capacity margin of zero provides a mark-up equal to the VOLL. For instance, where L and G_N^* are expressed in GW, for a VOLL of £10,000/MWh, $a = 10,000$ and $b = -1.123$, and for £2,000/MWh, $a = 2,000$ and $b = -1.101$.

For simplicity, linear variable costs are assumed for all generators. Fig. 4 shows an example of the price function given by (12); the curve behaves like a classical linear step-wise marginal cost supply function for small loads, but as the system approaches scarcity, the mark-up function becomes evident and soon becomes the dominate component of price.

The price mark-up function used is informed by a calibration exercise against GB market index price data (i.e., balancing mechanism prices) presented in [4]. Here the function is simplified to an exponential (Fig. 4). This is to keep the derivations presented in Section IV-C simple. However in practice any function of available capacity margin can be used (e.g., in [4] a hyperbolic and exponential are employed to provide a better match with the empirical price data). A sensitivity analyses on the impact on model results of alternative function parameters is briefly discussed in Section VI-B, with a deeper examination provided in [8].

Now the market price (12) no longer just depends on which generator is on the margin, it now depends on the overall margin, $G_N^* - L$ as well. Furthermore, since the price can exceed MC_n if n is on the margin, the question of whether a particular incremental MW of capacity within n is called upon or not must be considered because marginal generator n can still earn a positive gross margin.

Further, the expected gross margin (11) must be extended to consider the price function (12) and merit order operation. This is less straight-forward than under marginal cost-based pricing because the market price mark-up requires consideration of the total margin, M_N, as well as the marginal unit.

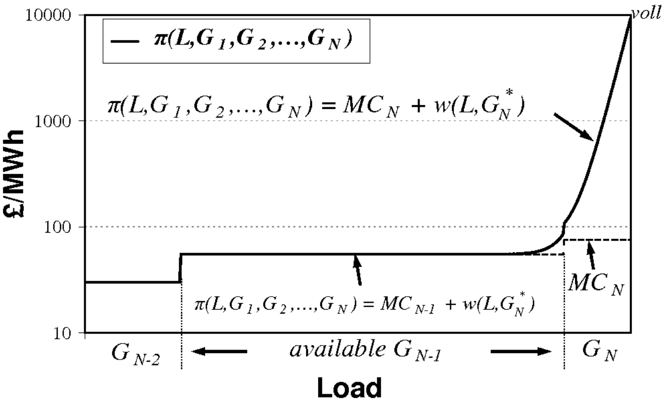


Fig. 4. Supply function for a given realised available capacity for generators $N-2$, $N-1$ and N with mark-up function defined by $w(L, G_N^*) = ae^{b(L-G_N^*)}$ (shown as black line). Marginal cost dashed, price is solid line.

By assuming that price mark-up is non-zero (i.e., $w(L, G_N^*) > 0$) only when generator N or $N-1$ is on the margin, calculation of the probability distribution of $w(L, G_N^*)$ can be achieved by considering the joint probability distribution of the capacity margins M_{N-1} and M_N.[9]

C. Expected price mark-up calculation

For each component of the MOND (4), we consider the joint distribution of capacity margins M_{N-1} and M_N, which is given by $f(M_{N-1}, M_N)$ with correlation $\sigma_{M_{N-1}}/\sigma_{M_N}$ [8]. The revenue that a MW belonging to generator $n \le N$ earns from price mark-up, RPM_n (£/MWh), is considered for $n \le N-2$, $n = N-1$ and $n = N$. The derivations for $n = N-1$ are included below, full derivations for the other cases can be found in [8]. If $n = N-1$, then RPM_n is broken down into two sub-cases. Firstly, if $M_{N-1} \le 0$ then

$$RPM_n = \int_{-\infty}^{0}\int_{0}^{v} (1-\rho_n) f(M_N, M_{N-1}) w(M_N) dM_N dM_{N-1}, \quad (13)$$

which is the case where all of generator $N-1$'s available capacity will be dispatched because load exceeds the available capacity of generators 1 through $N-1$. The inner integral upper bound is some value, v, above which the price mark-up is negligible owing to the large surplus margin. The integral lower bound is 0 because because $M_N \le 0$ results in zero price mark-up. Secondly, if $M_{N-1} > 0$ (which implies $M_N > 0$), then

$$RPM_n = \int_{0}^{v}\int_{M_{N-1}}^{v} p_{N-1}^{disp} f(M_N, M_{N-1}) w(M_N) dM_N dM_{N-1}, \quad (14)$$

where p_{N-1}^{disp} is the probability of dispatch of generator $N-1$, given by:

$$p_{N-1}^{disp} = \frac{(1-\rho_{N-1})M_{N-2}}{-1 \cdot (M_{N-1} - M_{N-2})} \quad (15)$$

[9] This assumption is reasonable in an aggregated capacity model with large generator type sizes.

where M_{N-2} is the surplus margin after all available generation lower in the merit order than $N-1$ has been dispatched.[10] M_{N-2} is a r.v. and computation of its pdf is awkward; however by approximating the realised value of G_{N-1} by its expectation: $E(G_{N-1}) = c_{N-1}(1-\rho_{N-1})$,[11] (15) can be expressed as a function of $N-1$ (see [8]).

Finally, by integrating over the subregions of the $\{M_{N-1}, M_N\}$ space (see [8]), the expected annual gross margin for generator n can be calculated as

$$GM_n = CGM_n + E[e_n] \cdot RPM_n. \tag{16}$$

This is an important extension to (11); by exploiting the properties of the probability distribution of capacity margins, this allows for the additional revenue received from market price mark-up to be calculated during the production costing process. To our knowledge this is the first time these derivations have been presented. The validity of these expressions have been confirmed in [8] by Monte Carlo simulation.

The MOND approximation technique is embedded within the dynamic investment model depicted in Fig. 1. More precisely, it is used to calculate the expected gross margin, GM_x^i, in (1) and realised gross margins in the wholesale electricity market. In both cases these are given by (11) or (16), depending on market bidding assumptions.

V. Case study assumptions

The new dynamic model is applied to an ‘energy-only’ market setting without a separate capacity market, with a initial capacity mix comparable to GB, a VOLL of £10,000/MWh, and a simulation time horizon of 30 years (2010-40). 100 Monte Carlo simulation runs are carried out for each plant type in each decision year. If used for an actual policy analysis, the impact that the number of Monte Carlo runs has on the model outcomes should be evaluated and, if possible, larger sample sizes used. The purpose of this paper is to present and illustrate a methodology, so extensive testing of the effect of the number of simulations was not undertaken. An indication of the importance of sample size is given by the standard error of the expected revenues. The samples are independent and identically distributed (i.i.d), so this error is proportional to $1/\sqrt{N}$ (here $N = 100$) [19]. For instance, for nuclear in 2020 in one run, the standard error was about 5% of the average revenue, which is reasonably precise. If the sample size was increased ten-fold to 1000 samples, one would expect an error of $5\%/\sqrt{10}$, or about 1.5%. Moreover, application of variance reduction methods (e.g., [20]) can result in smaller standard errors than i.i.d. sampling; however this is left for future research.

Fuel and carbon forward prices are based upon the UK Department of Energy and Climate Change (DECC) central case estimates [21]. Assuming investor price forecasts are similar to these estimates, the investor model uses the DECC estimate plus a r.v. to estimate future fuel prices. This r.v. is modelled as a mean reverting stochastic process.[12]

Investors assume a mean of 0% and standard deviation of 1% for annual load growth (cf. Section III-A). This is based on variations in demand growth [9] as well as the perception that economic growth could be offset by increased energy efficiency, thus allowing for small or even negative load growth. They also assume the parameter values for remaining build times (cf. Section III-A) are μ (mean) = one year and σ (standard deviation) = six months.

Investors in peaking capacity (i.e., OCGT) assume additional revenue can be obtained from the ancillary services (AS) market (here exogenously assumed to be £10,000/unforced MW/yr). Knowledge about revenues obtained from this market is somewhat uncertain in GB, therefore we have made an estimation consistent with the idea that AS revenues form a critical component of peaking capacity profitability, however they are insufficient by themselves to trigger investment. Similar models applied in the US (e.g., [9]) also do not treat the AS market explicitly, and instead assume a fixed (and relatively modest) contribution of AS to peaking unit gross margins. Furthermore these are likely to be a second order effect when considering generation investments on a decadal time scale and are likely to be relatively unimportant for cycling and base load capacity, given their relatively large capital costs.

The main ancillary services payment in GB is via the SO’s short-term operating reserve (STOR) market. Units who participate in this market can chose to participate in both the STOR market and the energy-only market. Under a “flexible” contract, units specify availability windows, outside of which they can participate in the energy market. Alternatively, under a “committed” contract they participate in only the STOR market. The revenues from this market are relatively small (e.g., typical availability payments are around £5/MW/hr [22]) and because of their small size, are in reality only considered for low capital cost plant investment (i.e., peaking units).

Hourly wind production is derived by scaling total installed wind capacity by simulated hourly aggregated GB wind capacity factors (CFs). Onshore, the methodology described in [23] is applied, whereby hourly wind speed data for 2005-09 is obtained (to match the empirical load data), and aggregate GB CFs are obtained by applying regional weightings derived from the current wind capacity in operation, construction, or consented. For offshore wind, we instead use outputs from a validated (see [24]) fully compressible, nonhydrostatic mesoscale atmospheric 3km grid point model. The main justifications for using wind speed measurement data when simulating aggregate onshore wind production are: 1) there is the lack of extensive time series data with adequate temporal resolution for currently operational farms in GB; and 2) the amount and geographical dispersion of onshore wind farms will grow over time, and will include production from regions where CFs and correlations with loads and wind from other areas are not currently known.

The total installed wind capacity is exogenous to the model,

[10] The -1 scalar is applied to the denominator of (15) in account of M_{N-2} being negative. If it was positive, then G_{N-1} would not be dispatched (i.e., $p_{N-1}^{disp} = 0$), which is not considered in (14).

[11] Note here that c_{N-1} is the capacity of the generator type $N-1$, which is the sum of a number of individual units who share the same capacity and FOR characteristics (see section IV-A).

[12] See [8] for parameters used.

and is expected to increase linearly from 2010 levels (approximately 2 GW or 3% of total installed capacity), up to 30 GW by 2020 with a maximum of 35 GW in 2025, after which it levels off. We justify this approach by the fact that, to date, large-scale investment in wind capacity is driven by policy, rather than economic considerations. It is therefore assumed that policies promoting wind investment are successful in meeting renewables targets,[13] and the purpose of this work is to provide insights into the response of investment in thermal generation and subsequent levels of security of supply risk. This magnitude of increase in total installed wind capacity in a system with around 75 GW of total transmission connected capacity and a maximum and minimum annual demand of 60 GW and 20 GW, respectively, transforms the GB system from a low to high wind penetration. Note that a positive probability of wind exceeding load (negative net demand) arises only for scenarios in which the wind penetration exceedes 32 GW. Furthermore the fraction of hours in which this occurs diminishes as load growth continues after 2025. For example, for a wind penetration of 35 GW (the maximum simulated level) in 2025, of the 43,825 simulated wind and empirical 2005-09 demand hours, 113 experienced negative net demand (0.3% of hours).

The residual load facing thermal units for a particular hour is calculated by subtracting the hourly wind production from the full load. Fig. 5 shows an example of the impact on the 2005-09 residual load histograms as penetration of wind increases.

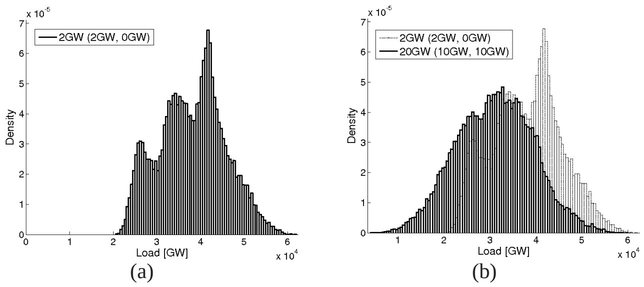


Fig. 5. Result of increasing installed wind capacity from (a) 2 GW to (b) 20 GW on residual load histograms. Data shown is for 2005-09. Numbers in brackets indicate volume of onshore and offshore capacity respectively.

Data on initial 2010 system capacity in Table I is derived by aggregating GB capacity data [26] into the five capacity types. Unit sizes are 500 MW for nuclear and coal, 200 MW for CCGT and 50 MW for OCGT. To keep the model simple, minor sources of peaking capacity such as oil and pumped storage is combined with OCGT. CHP and hydro are aggregated with CCGT plant to obtain the unit totals shown in Table I. This table also shows the financial and technology input assumptions, including capacity cost assumptions, total annualised fixed costs ($TAFC_x$) and total interest accumulated during construction ($TIAC_x$).

We assume there will be no load growth until 2020 (although, as explained, realised growth varies around the mean rate). This is broadly in line with central Updated Energy Projections published by DECC [27] and base forecast winter peak demand figures from the GB System Operator (SO) [26]. Expected electricity demand after this point is assumed to grow at 1% per year.

VI. Case Study Results

A. Base case results

The model has been implemented using in the Matlab/Simulink environment. The computational efficiency of the MOND technique allowed for each production costing run to execute in under 1.5 seconds.

Fig. 6 shows the evolution of total installed capacity in the simulation. Also shown is the full and de-rated capacity margin. The de-rated margin is the ratio of de-rated capacity (DC) (installed capacity scaled by expected availability at peak demand) to most probable peak load (PL); i.e., $[DC]/[PL]-1$. The FORs in Table I are used to de-rate conventional capacity, and for wind the long-term capacity credit (%) values are calculated using only those hours within 10% of peak demand [23]. These range from 9-35% depending on level of installed capacity (the higher the total installed capacity, the lower the capacity credit). The forecast for peak demand is obtained from the 99.9% percentile of the year's MOND cdf for full load (i.e., the load which is exceeded approximately 9 hours per year). The use of de-rated margin is preferable when calculation of an absolute level of risk is difficult. Moreover it can easily be compared with the GB SO's current estimate of what constitutes an acceptable margin.[14]

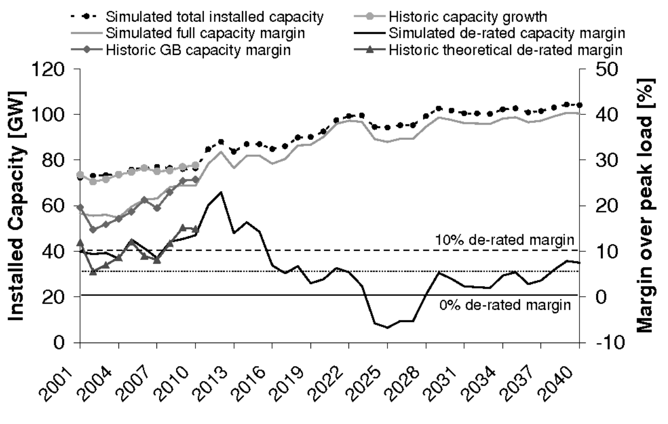


Fig. 6. Plot of simulated capacity growth, de-rated and full capacity margins. The "historical theoretical de-rated margin" (2001-10) is what the forecast de-rated margin would have been for a given winter using the GB SO's assumptions about plant availability and winter peak [28].

To compare the performance of the simulation against historic trends in GB, the model was run from 2001-10 and a comparison between the modelled and actual capacity margins was performed.[15] The comparison shows that simulated margins do not perfectly match historic trends in all years (e.g., 2002/3), however there is a reasonably good agreement of our model with reality, which gives a degree of confidence

[13]For instance, the UK Government has a target of around 30% renewable electricity generation by 2020 in order to meet the binding European Union target for renewable energy [25].

[14]For example, see Appendix to [28] and the UK Government's recent consultation on GB electricity market reform [25].

[15]See [4] for cost and initial capacity data.

TABLE I
GENERATOR INPUT ASSUMPTIONS WITH SYMBOLS DEFINED IN SECTION III.

Technology x	Therm. eff.	ρ (FOR)	Capex £/kW	FC £/kW/yr	Var. O&M £/MWh	Lifetime α (yrs)	Build τ (yrs)	WACC r (real)[a]	$TAFC$ £/unfor.MW/yr	$TIAC$ £/MW	Initial (GW)	No. units
Nuclear	0.36	0.10[b]	2,913	37.5	1.8	40	7	0.09	400,750	931,170	11	22
Coal	0.35	0.14	1,789	38.0	2.0	40	5	0.07	216,710	344,100	27.5	55
CCGT	0.53	0.13	718	15.0	2.2	25	3	0.07	91,840	96,030	28.6	143
OCGT	0.39	0.10	359	15.0	4.4	40	2	0.07	47,250	36,690	7.7	154

[a]Assuming a 2.5% rate of inflation. [b]Recent years have shown a decline in the annual availability of the GB nuclear fleet (likely due to age), therefore this value is reduced to 75% for existing nuclear capacity. New nuclear builds are expected to have 90% availability.

in the realism of our future projections. The average absolute difference between the historical theoretical de-rated margin and the simulation was 1.6% with a standard deviation of 1.3%. A comparison of available historical data on CCGT expansion during 2004-09 [29] showed that CCGT investments triggered by the simulation (3.9 GW during 2004-09) did not correspond in all years, however the volumes and timing were not unreasonably different.[16] A further analysis of differences in levels of actual and simulated total installed capacity can be found in [8].

The future trend shows an erosion of de-rated capacity margins after around 2015. This coincides with the Large Combustion Plant Directive[17] plant retirements and rapid offshore wind growth. Of the 30 simulated future years, the average de-rated margin is 5.6% with a standard deviation of 7.1%. De-rated margins are negative in 4 years, below 5% in 15 years, and below 10% in 25 years. For those years where margins are below 10%, an average shortfall of 1 GW of capacity was projected. The GB SO has recently stipulated that a de-rated capacity margin of 5 GW over expected peak demand is desirable (see Appendix to [28]). These simulation results suggest that a lower than desirable level of adequacy risk could potentially occur.

The annual LOLE and EEU was also calculated (cf. Section IV-A). The average annual LOLE across the 30-year simulation was 0.03 hrs/yr with a standard deviation of 0.05, and average annual EEU of 5.7 GWh (less than 0.002% of typical year's total annual energy demand). The yearly LOLE together with the volume of hypothetical additional capacity required to meet a 5 GW de-rated capacity margin at peak is plotted in Fig. 7. A value of zero implies that de-rated margin is in excess of 5 GW.

These projected risk and de-rated capacity margin figures suggest that the system may experience tight supply conditions during peak demand in some years. Some of these results can perhaps be explained by inspection of the residual load histograms from Fig. 5(b); the shape of the right-most tail suggests that even with very high penetrations, wind power does not contribute in all high demands periods. However the frequency of these high-demand/low-wind periods is too low to justify investment by private investors. And it is these very high-demand hours when the potential for a capacity

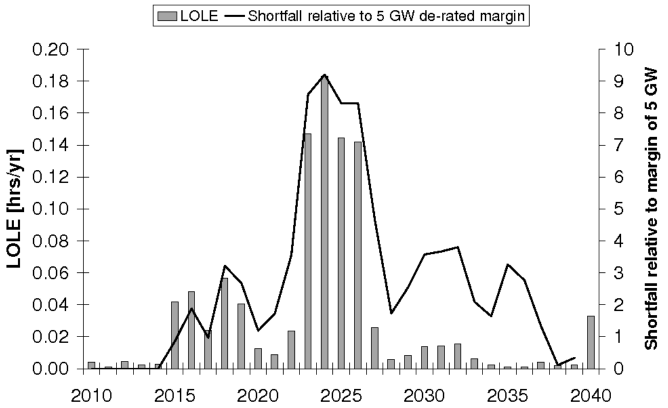


Fig. 7. Plot of simulated LOLE (bars) and capacity shortfall over 5 GW de-rated capacity margin (solid line).

shortfall is highest (excluding here SO actions such as voltage reductions). From a policy perspective, it is arguably uneconomical to design policies aimed at ensuring there is adequate generating resource available for these low-frequency events; an alternative approach would be to encourage demand-side participation through smart grids and smart metering.

Further, an analysis of generator revenues showed symptoms of a boom and bust investment cycle. Simulated OCGT total gross margins were positive in only 7 of the 30 simulated years. The largest investment years were 2017/18, which included the forecast prices for 2023-25, the period when gross margins are highest. However investment reduces in 2019-22 in response to expectations about prices being dampened (although not sufficiently to prevent an overshoot) as new investment in OCGT and other technologies enter the system. The pattern of high gross margins directly corresponds to those years where adequacy risk is highest (cf. Fig. 7). The boom in OCGT investment in expectation of the high gross margins after 2023 is followed by a bust phase around 2026 when large volumes of new nuclear capacity begin entering the system. This increases the capacity margin but reduces profitability for peaking units.

B. Sensitivity analyses

In order to test the robustness of the model, sensitivity analyses have been performed on a number of model assumptions. These included:

1) The ability of plant owners to exercise market power and volumes of revenues received by doing so; this is

[16]CCGT investment was the primary source of capacity expansion in GB dring this period.

[17]A control on emissions from heavily polluting large combustion plant introduced by the European Union in 2001. Approximately 11 GW of emission-intensive capacity is expected to be decommissioned by 2016 under this legislation

achieved by calculating expected gross margins using (11) instead of (16).

2) The impact on investment of the expected scarcity price; this is achieved by reducing the VOLL and recalibrating price mark-up (see section IV-B) when calculating (16).
3) Altering investor estimates about construction lead times to match reality (i.e., as shown in Table I) in order to test the hypothesis that differences between realised and investor expectations about prices is due to the uncertainty surrounding capacity under construction.

Also the aggregate investment response (2), investor risk preferences, uncertainty about load growth and varying economics of peaking plants (revenue from AS and fixed costs) were all found to affect investment dynamics.

Overall, the model's qualitative behaviour was reasonable for these sensitivities and provided some useful insights, particularly when comparing the oligopolistic base case to the perfectly competitive market results. A pattern of increased relative levels of risk and erosion of de-rated capacity margins was experienced during the 2020s to some degree in all cases. Furthermore, the period of highest security of supply risk is 2023-28, however the magnitude of this risk differs between experiments. A prolonged period of increased security of supply risk is experienced throughout the 2020s for the perfectly competitive market case. Plainly these dynamics may differ in a model representing more sophisticated firms who account for the effect of their investments upon mark-ups for their existing fleet. In this case, firms may deliberately not invest in order to keep prices (including mark-ups) high. Also, providing investors with perfect foresight about capacity under construction produces less investment. This leads to more frequent periods of relatively high LOLE and low de-rated margins after 2020. However, generators experience positive gross margins in more years due to lower surplus margins, and hence higher prices.

VII. Conclusion

We have presented the MOND technique for calculating expected output, costs and revenues of thermal generation subject to varying load and random independent thermal outages. This method has been adapted for use in a dynamic capacity market model with high penetrations of wind by performing a residual load calculation with simulated wind outputs. An 'energy-only' market setting has been used to estimate the economic profitability of capacity investments. Using relative levels of de-rated capacity margin and LOLE as risk metrics, simulation results for GB show that levels of generation investment lead to a mild increase in generation adequacy risk in some years, with erosion of de-rated capacity margins in the mid 2020s, and very tight supply conditions are experienced during a small number of peak hours. Many new investments, particularly peaking units, were unable to recover their fixed costs.

If the model presented here were to be used for an actual policy analysis, a number of key areas of future research are recommended. These include: 1) a survey of investors regarding their use of analytical approaches when evaluating investment decisions under uncertainty and, if appropriate, modification of investor risk-averse decision making, 2) to develop an explicit representation of the heterogeneity of investor response, and 3) modelling endogenous bilateral contracting.

An area of possible future research would be to extend the MOND technique to a two-area (e.g., Scotland-England) system with a transmission constraint. This could then be used to assess the impact of congestion on wind power availability and thermal generation investment in each area. This could be implemented using a Mix of Bivariate Normals reflecting the correlations in loads and wind in each area (e.g., as in [2]), based on two area production costing methods (e.g., see [30]). To maintain computational efficiency, the model could draw upon scenario reduction schemes (e.g., see Chapter 7 of [31]) when applying multi-area production costing and making an assessment of curtailment of wind production due to transmission congestion.

The next stage of this research will address whether explicit capacity mechanisms such as tendering for strategic reserve (e.g., [25]) and capacity markets (e.g., [9]) can be designed to alleviate resource shortfall and prevent investment overshoot. The results here indicate that such a mechanism may be desirable to improve reserve margins in the mid 2020s in Great Britain.

Acknowledgment

The authors acknowledge the valuable input from C. Dent, G.P. Harrison, T. Johnson and S. Hawkins. Also thanks to the Electricity Policy Research Group at the University of Cambridge for facilitating this collaboration. This research formed part of the UKERC programme and was supported by the UK Research Councils under NERC award NE/C513169/1. Partial support for B.F. Hobbs was provided by the U.K. EPSRC Supergen Flexnet program and the U.S. National Science Foundation through EFRI Grant 0835879.

References

[1] G. Gross, N. V. Garapic, and B. McNutt, "The Mixture of Normals Approximation Technique for Equivalent Load Duration Curves," *IEEE Trans. Power Systems*, vol. 3, no. 2, pp. 368–374, May 1988.

[2] J. G. Iñón, "Long-run Probabilistic Production Costing-Based Analysis of Electricity Market Designs for Long-Run Generation Investment Incentives under Spot-Pricing and Transmission Open-Access," Ph.D. dissertation, John Hopkins University, 2012, in preparation.

[3] B. F. Hobbs, J. G. Iñón, and S. E. Stoft, "Installed Capacity Requirements and Price Caps: Oil on the Water, or Fuel on the Fire?" *The Electricity Journal*, vol. 14, no. 6, pp. 23 – 34, 2001.

[4] D. Eager, J. W. Bialek, and T. Johnson, "Validation of a Dynamic Control Model to Simulate Investment Cycles in Electricity Generating Capacity," in *Power and Energy Society General Meeting, 2010 IEEE*, July 2010, pp. 1–8.

[5] "Waiting for the Boom: a Simulation Study of Power Plant Construction in California," *Energy Policy*, vol. 29, no. 11, pp. 847 – 869, 2001.

[6] F. Olsina, F. Garces, and H.-J. Haubrich, "Modeling Long-Term Dynamics of Electricity Markets," *Energy Policy*, pp. 1411–1433, 2006.

[7] B. F. Hobbs and Y. Ji, "A Bounding Approach to Multiarea Probabilistic Production Costing," *IEEE Trans. Power Systems*, vol. 10, no. 2, pp. 853–859, May 1995.

[8] D. Eager, B. F. Hobbs, and J. W. Bialek, "Dynamic Long-Term Modelling of Generation Capacity Investment and Capacity Margins: a GB Market Case Study," *EPRG Working Paper Series*, no. 1201, Jan. 2012, www.eprg.group.cam.ac.uk/category/publications/.

[9] B. F. Hobbs, M.-C. Hu, J. G. Iñón, S. E. Stoft, and M. P. Bhavaraju, "A Dynamic Analysis of a Demand Curve-Based Capacity Market Proposal: The PJM Reliability Pricing Model," *IEEE Trans. Power Systems*, vol. 22, no. 1, pp. 3 –14, Feb. 2007.

[10] "UK Electricity Generation Costs Update," Mott MacDonald, Tech. Rep., June 2010, www.decc.gov.uk/en/content/cms/about/ec_social_res/analytic_projs/gen_costs/gen_costs.aspx.
[11] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent Measures of Risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203 – 228, 1999.
[12] B. Willems and J. Morbee, "Market Completeness: How Options Affect Hedging and Investments in the Electricity Sector," *Energy Economics*, vol. 32, no. 4, pp. 786 – 795, 2010.
[13] R. Billinton and R. N. Allan, *Reliability Evaluation of Power Systems. 2nd ed.* Plenum, 1994.
[14] J. Endrenyi, *Reliability Modeling in Electric Power Systems*. Wiley, 1978.
[15] Ofgem, "Transmission Access Review," www.ofgem.gov.uk/Networks/Trans/ElecTransPolicy/tar/Pages/Traccrw.aspx.
[16] Pöyry, "Impact of Intermittency: How Wind Variability Could Change the Shape of the British and Irish Electricity Markets," July 2009, www.poyry.com/linked/group/study.
[17] R. Sioshansi and S. Oren, "How Good are Supply Function Equilibrium Models: an Empirical Analysis of the ERCOT Balancing Market," *Journal of Regulatory Economics*, vol. 31, pp. 1–35, 2007.
[18] J. B. Bushnell, E. T. Mansur, and C. Saravia, "Vertical Arrangements, Market Structure, and Competition: An Analysis of Restructured US Electricity Markets ," *American Economic Review*, vol. 98, no. 1, pp. 237–66, 2008.
[19] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*. McGraw-Hill, 2001.
[20] A. Breipohl, F. N. Lee, J. Huang, and Q. Feng, "Sample Size Reduction in Stochastic Production Simulation," *IEEE Trans. Power Systems*, Aug. 1990.
[21] Department of Energy and Climate Change, "Communication on DECC Fossil Fuel Price Assumptions," www.decc.gov.uk/en/content/cms/about/ec_social_res/analytic_projs/ff_prices/ff_prices.aspx.
[22] National Grid, "Short-Term Operating Reserve," www.nationalgrid.com/uk/Electricity/Balancing/services/STOR/.
[23] P. E. O. Aguirre, C. J. Dent, G. P. Harrison, and J. W. Bialek, "Realistic Calculation of Wind Generation Capacity Credits," in *Integration of Wide-Scale Renewable Resources Into the Power Delivery System, 2009 CIGRE/IEEE PES Joint Symposium*, July 2009, pp. 1–8.
[24] S. Hawkins, D. Eager, and G. P. Harrison, "Characterising the Reliability of Production from Future British Offshore Wind Fleets," in *Renewable Power Generation Conference, IET*, Sept. 2011.
[25] Department of Energy and Climate Change, "Electricity Market Reform Consultation Document," Tech. Rep., Dec. 2010, www.decc.gov.uk/en/content/cms/consultations/emr/emr.aspx.
[26] "Seven Year Statements," National Grid, Tech. Rep., www.nationalgrid.com/uk/Electricity/SYS/.
[27] Department of Energy and Climate Change, "Energy and Emissions Projections," www.decc.gov.uk/en/content/cms/about/ec_social_res/analytic_projs/en_emis_projs/en_emis_projs.aspx.
[28] "Winter Outlook Consultation 2010/11," National Grid, Tech. Rep., www.nationalgrid.com/uk/Gas/TYS/outlook/.
[29] Digest of United Kingdom Energy Statistics, "Chapter 5: Electricity, Tables 5.6 and 5.7," www.decc.gov.uk/en/content/cms/statistics/publications/dukes/dukes.aspx.
[30] L. R. Noyes, "Two-Area Probabilistic Production Costing by the Method of Bi-Variant Cumulants," *IEEE Trans. Power Apparatus and Systems*, vol. 102, no. 2, pp. 433–443, Feb. 1983.
[31] D. Burke, "Accommodating Wind Energy Characteristics in Power Transmission Planning Applications," Ph.D. dissertation, University College Dublin, 2010.

Dan Eager (SM'08) received the B.Sc. degree in mathematics and computer science from The University of Sussex, UK, in 2004, and the M.Sc. degree in operational research from The University of Edinburgh in 2007. He is currently undertaking a Ph.D. within The University of Edinburgh's Institute for Energy Systems under the supervision of G.P. Harrison, J.W. Bialek and T. Johnson.

Benjamin F. Hobbs (F'08) received the Ph.D. degree in environmental system engineering from Cornell University, Ithaca, NY, in 1983. He is Schad Professor of Environmental Management at the Johns Hopkins University, Baltimore, MD. In 2009-2010 he was on sabbatical at the Electricity Policy Research Group at the University of Cambridge. He chairs the California ISO Market Surveillance Committee and is Director of the JHU Environment, Energy, Sustainability & Health Institute.

Janusz W. Bialek (F'10) is Professor of Electrical Power and Control with the School of Engineering and Computing Sciences at Durham University, UK. He holds M.Eng (1977) and Ph.D. (1981) degrees from Warsaw University of Technology, Poland. In 1981- 1989 he was with Warsaw University of Technology, in 1989-2002 with Durham University, UK, and in 2003-2009 with The University of Edinburgh. He has co-authored 2 books and over 100 technical papers.

CHARACTERISING THE RELIABILITY OF PRODUCTION FROM FUTURE BRITISH OFFSHORE WIND FLEETS

*S. Hawkins, D. Eager, G.P. Harrison**

*Institute for Energy Systems, School of Engineering, University of Edinburgh; *Gareth.Harrison@ed.ac.uk*

Keywords: Capacity value, offshore wind generation, mesoscale modelling.

Abstract

The extent to which large volumes of offshore wind can contribute to a secure and reliable electricity supply is a subject of much debate. Key to providing credible answers requires a detailed understanding of the wind resource and its variability in time and space. Here, a mesoscale atmospheric model was employed to create a ten year hindcast of British onshore and offshore wind speeds. This was used to simulate the output of a British offshore wind fleet and combined with demand data to assess reliability during periods of high demand. Further, capacity value calculations using Effective Load Carrying Capability for the combined onshore and offshore GB wind fleet provides an estimate of the long-term reliability of production.

1 Introduction

Integrating large amounts of variable renewable generation into the electricity network presents a significant challenge and is the subject of much debate. This is particularly true in the UK where wind generation is expected to become a significant supplier of energy, with a large increases in capacity expected offshore, perhaps in excess of 30GW by 2030, up from around 1GW today.

Debate centres on the question: 'to what extent can a variable and stochastic resource contribute to a secure and reliable electricity supply?' Key to answering this is a detailed understanding of the wind resource and its variability in time and space. However, there are relatively few sources of offshore observations with sufficient temporal resolution or accuracy to address this. In a future system with high penetrations of wind, the temporal variability of wind will determine numerous characteristics such as the capacity value of wind and the amount of reserve required to maintain an adequate level of system security.

This paper presents the results of a high resolution re-analysis using a mesoscale atmospheric model to recreate ten years of hourly wind speeds across Great Britain (GB) and surrounding waters. Wind speeds are extensively validated against observations from a number of available buoys, lightships and offshore platforms. The dataset is used to simulate ten years of wind production. Taking inspiration from capacity value calculations for onshore wind [1] this new data is used to produce the first credible estimate of the contribution of British offshore wind generation in supporting demand.

2 Mesoscale modelling

2.1 Simulation

Mesoscale atmospheric modelling is becoming widely used in the wind energy field, both for short-term forecasting and longer-term resource assessment. Mesoscale models are computationally demanding, so many studies either simulate relatively short time periods or employ statistical downscaling to reduce the computational requirement needed to capture a representative period. However, short term analyses do not fully capture wind speed variability, while statistical approaches do not produce continuous historic time-series which can be matched with historic patterns of energy demand.

This study uses the Weather Research and Forecast (WRF) model [7], a fully-compressible non-hydrostatic mesoscale model with multiple boundary-layer, land surface, microphysics and cloud physics options. The model was configured with three nested domains down to 3km resolution (Figure 1). Boundary conditions were taken every six hours from the NCEP Global Forecast System Final Analysis dataset at 1^{o} resolution. Two- way nesting and analysis nudging was used on all domains. The main physics options are summarised in Table 1.

Ten years were simulated, from 2001-2010 inclusive, on the UK Research Council's high performance computing platform, HECToR.

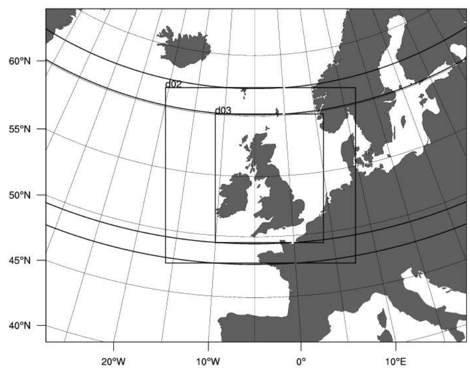


Figure 1: Nested domains at 27, 9 and 3km resolution

Domain	1	2	3
Resolution (km)	27	9	3
Integration timestep (s)	135	45	15
Analysis nudging	y	y	y
Cumulus scheme	Kain-Fritsch		None
PBL scheme	MYJ [4]		
Surface layer	Monin-Obhukhov [5]		
Land surface scheme	NOAH		
Land use dataset	MODIS		

Table 1. Summary of mesoscale model configuration.

WRF uses a terrain-following, pressure based vertical coordinate system. The vertical resolution was increased close to the ground to reduce any errors associated with interpolation to fixed heights. Wind speeds were interpolated to hub height from the closest model level:

$$U_z = U_m \cdot \ln(z / z_0) / \ln(z_m / z_0) \quad (1)$$

where U_z is the wind speed at hub height z. U_m is wind speed at the closest model level z_m and z_0 is the local roughness length taken from WRF. Over water, WRF uses a Charnock formulation for roughness length.

2.2 Wind farm load factor

For existing (as well as planned and under construction) onshore and offshore wind farms, hourly wind speeds were converted to power output using the manufacturer's power curve for the appropriate turbine. The make, model and size of turbine are specified in the RenewableUK wind farm database.

The location of future offshore wind farms were taken from the Crown Estate leasing rounds. A generic 3MW turbine was assumed for Round 2 sites, and a generic 5MW turbine for Round 3, based on commercially available models. The final installed capacity in each offshore site was assumed to be distributed in proportion to the maximum lease capacities.

Overall GB-level aggregate load factors (LFs) were computed as averages weighted by final installed capacity. That is, if LF_t^n represents the LF of wind farm n at time t, then the aggregate LF at time t is calculated as:

$$LF_t = \sum_{n=1}^{N} (LF_t^n \cdot C^n) / \sum_{n=1}^{N} C^n \quad (2)$$

where C^n is the final installed capacity of wind farm n. Aggregate offshore and onshore LF are calculated separately, i.e. n is restricted to either offshore or onshore farms. For offshore farms, this means longer term calculations are weighted towards the larger Round 3 sites.

2.3 Validation

Wind speeds were validated against onshore met stations and offshore buoys, lightships and platforms. Standard error statistics of Bias (B), Mean Percentage Error (MPE), Root-Mean-Square Difference (RMSD) and coefficient of determination (R^2) were calculated.

	n	B m/s	MPE %	RMSD m/s	R^2
Met stations	220	0.02	-0.5	0.44	0.96
Buoys	9	0.24	6.25	1.16	0.82
Lightships	4	-0.38	-1.99	1.30	0.91
Platforms	3	0.30	3.48	1.54	0.93

Table 2. Summary of error statistics by observation type

Table 2 summarises the error statistics by class of observation. The performance is generally good, with high correlation values. Onshore the agreement between simulated and observed wind speeds was very good, with overall high correlation and low bias. Offshore a seasonal bias of -1 to -2ms was found in summer months. This merits further investigation, however it does not affect the capacity value analysis presented here which is based only on winter wind speeds.

Monthly measured LFs for the largest onshore wind farms in each region of the UK (covering 196 wind farms with a totalled installed capacity of 2.7GW) were compiled from Ofgem's Renewable Obligation Certificate (ROC) Register for the period from April 2006 to December 2010. LFs for existing offshore wind farms were also compiled from the time they became operational until December 2010. By the end of that period, that amounted to 8 wind farms with a total installed capacity of around 1 GW.

Simulated LFs at the same sites were derived from the modelled hourly wind speeds and then averaged to monthly values. No accounting for wake losses or other array losses was carried out at this stage. Figures 2 and 3 show the agreement between average monthly LF averaged across onshore and offshore sites. Onshore, the predicted LFs were found to be consistently higher than observed. This would be expected before losses have been considered. A linear scaling factor of 0.69 gave the best adjustment between simulated and observed LFs (R^2=0.94).

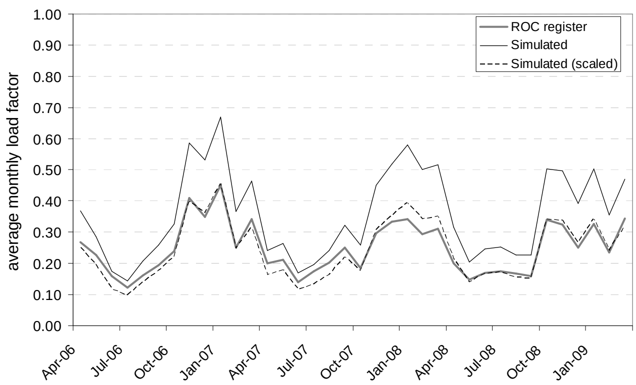


Figure 2: Average monthly LF for existing onshore wind farms. The thick grey line shows the actual weighted average LF from 196 large wind farms. The solid black line shows simulated LFs at 100% availability/no losses and the dashed line shows these values scaled by 0.69.

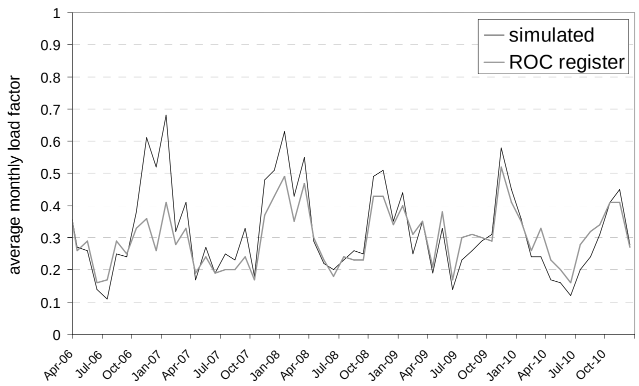


Figure 3: Average monthly LFs for existing offshore wind farms. The thick grey line shows weighted average LF for all operational offshore The solid black line shows simulated LFs with no adjustment for losses. The large deviation in the first winter period is due to low technical availability in the early stages of some offshore farms.

Offshore, the predicted LFs were already in quite close agreement to the observed values, except for two winter periods were technical availability at a number of offshore wind farms was very low. LFs were slightly too low in summer, confirming the seasonal wind speed bias seen at observation sites. For this reason, no further accounting for wake losses or technical availability was performed for offshore sites.

3 Reliability analysis

3.1 Historic wind and demand time series

The relationship between wind generation and electrical demand is of primary interest when analysing the reliability of the wind resource. Of particular interest is the availability of the wind resource during periods of high demand. In Britain the highest demands are driven by low temperatures occurring during winter (November-March). It is during these periods that the adequacy risk is typically highest.

Historic aggregated half-hourly demand data going back to April 2001 is available from the GB System Operator (SO), National Grid [9]. The GB 'IO14_DEM' data is the most applicable for generation adequacy calculations because this is based on operational generation metering and includes station load and pumped storage (PS) pumping [1]. However this data is inconsistent as prior to April 2005 it relates to England and Wales only. The 'INDO' demand measure, which excludes station load and PS pumping, is available for the entire period. The winter 'IO14_DEM' values can be approximated by raising the 'INDO' measure by 600 MW and is used where the 'IO14_DEM' data is not available.

To account for underlying changes in absolute levels of peak demand, each winter's demand is normalised by out-turn "Average Cold Spell" (ACS) peak demand and rescaled to 60 GW. ACS peak demand is forecast each year in advance of the forthcoming winter by the SO, and is described as having "a 50% chance of being exceeded as a result of weather variation alone" [8]. The out-turn ACS peak is calculated post winter and is a measure of what peak demand would be given a winter's underlying demand patterns and "typical" winter peak weather conditions [8]. This makes it suitable value for the normalisation. These values can be found in [1] and [10].

The half-hourly data is transformed to hourly resolution by taking the hourly demand to be the maximum of the two half-hour periods. Finally, the normalised hourly demand data is then matched with the hourly simulated wind LFs. The time series spans 9.5 consecutive winters from winter 2001/02 to December 2010, totalling 34,128 demand hours.

Figure 4 shows the simulated average aggregate long-term LFs for wind generation during the highest demand hours. The 90%+ normalised demand hours are categorised into 1% bins and the cumulative number of hours at each demand level are indicated on the graph (i.e., each label indicates the number of hours demand is at or above *x*). Note that demand levels above 100% of peak are possible on account of ACS peak being exceeded in some years. Interestingly, the pattern of average LFs shows a good agreement with the analyses of transmission metered wind farms presented in [14]. However absolute levels of average LF are higher with around a 45% LF at 90% to 95% levels of demand compared to 20% in [14]. This is hardly surprising given the increased geographic diversity of the wind capacity. As mentioned earlier, no scaling factors were applied to the simulated offshore LFs, which may lead to systematic over-estimation in the results. Therefore a second case where weighted offshore LFs are scaled by 0.69 is also included in Figure 4. This reduces the average LF at 90% to 95% demand levels to around 35%. The level of deterioration in average LFs at high demands is less severe than in [14], although this is based on just 2 hours of simulated data. It highlights the challenge for determining the availability of wind at high demand levels [14].

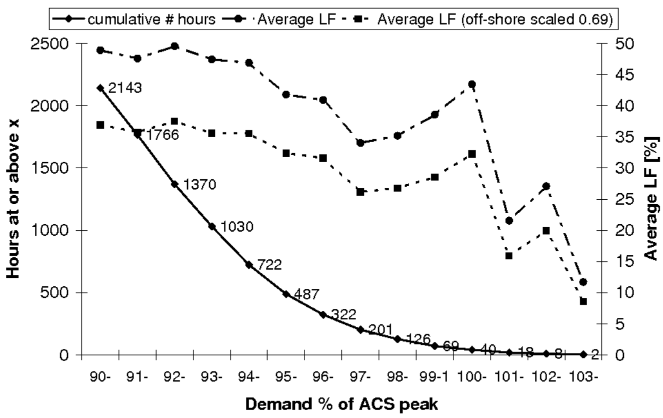


Figure 4. Average long-term LFs for highest demand levels (right y-axis): base-case and wind scaled by 0.69 (upper and lower dashed lines). The demand hours per 1% normalised demand bin are shown on the left y-axis.

3.2 Risk metrics

The loss-of-load probability (LOLP) in a particular period is defined as the probability that available generation is unable to meet demand:

$$LOLP = p(X < D), \tag{3}$$

where X is the available generation and D is the system demand, both of which are random variables. The loss-of-load expectation (LOLE) is the expected number of periods over a given period in which available generation is unable to meet demand. So for a given time horizon:

$$LOLE = \sum_{t}^{T} LOLP_t, \tag{4}$$

where $LOLP_t$ is the LOLP in period t. Here, the period t is assumed to be one hour and T spans a number of years.

3.3 Capacity value calculations

The use of capacity value is common when measuring the contribution of renewable energy generation to meeting demand. Here the Effective Load Carrying Capability (ELCC) defines the capacity value (or capacity credit). The ELCC for a particular level of additional generating capacity estimates the amount of additional demand that can be served due to the extra generation whilst maintaining the original level of system risk [6]. The purpose of this study is to demonstrate an application of the mesoscale model, not to determine an absolute measure of wind generation capacity value for GB. Moreover, a specific methodology has been applied to calculate capacity value, and while the authors concede that this is not necessarily cutting edge probability theory (e.g., see [14]) it is a fully valid contribution to defining current approximations of capacity value in GB (e.g., [11,12]). What is more, this is understood to be the first time that capacity value for a combined on- and offshore wind resource has been calculated using the ELCC approach.

The capacity value is estimated as follows. Consider some additional wind generation w which increases overall system capacity. If the system LOLP in hour t before the additional generation is added is the “initial” LOLP, then adding the additional generation will reduce the LOLP. The total reduction depends on the reliability of the additional generation. This reduced LOLP can be expressed by:

$$LOLP^{*} = p(X < D - W), \tag{5}$$

where W is the contribution to demand from the additional generation. Similarly, using the same principle as (4), the reduced 9.5 winter LOLE can also be determined.

The ELCC for the additional generation is found by increasing demand until the reduced LOLP* risk returns to its original value. Here the interest is in the ELCC across the entire time horizon; however calculation of the ELCC for a single period is the starting point. This is given by [14]:

$$p(X < D) = p(X < D + d_{ELCC} - W), \tag{6}$$

where d_{ELCC} is the ELCC. This can be extended over multiple periods to:

$$\sum_{t}^{T} p(X_t < D_t) = \sum_{t}^{T} p(X_t < D_t + s_t d_{ELCC} - W_t), \tag{7}$$

where s_t is a scalar applied to the ELCC in order to account for the level of demand being experienced. Or put another way, the scalar places a higher weight on the highest demand periods when solving (7).

3.3.1 Treatment of conventional generation

The next step is to construct a probability distribution for available conventional generation. Here, the term conventional generation covers all forms of generation currently connected to the high voltage transmission system in GB, with the exception of wind. Furthermore, the availability of conventional generation is assumed to be independent of demand and available wind capacity. Technical plant availability data is not available in GB. However most generating companies try to make available as much capacity as possible at time of highest demand (to not forgo high wholesale market prices), thus availability is a function of the unit’s forced outage rate (FOR), which it is reasonable to assume are independent [1].

Generation unit data is taken from the National Grid Seven Year Statement [10] and the expected winter peak availabilities in their Winter Outlook [11] are used as FORs. This data is summarised in Table 3. The Unit Effective Capacity (UEC) in [10] has been used for all units, apart from those which are transmission constrained; in this case the individual UEC is scaled in order to match the transmission limit. Hydro units belonging to the same hydro scheme are combined into single pseudo-units owing to their resource interdependence.

Power station type	*No. units*	*Capacity (GW)*	*Assumed availability*
Nuclear	22	10.1	0.75
Interconnector	1	2	1.00
Hydro	9	1.1	0.60
Coal	62	27.9	0.90
Oil	4	2.7	0.80
Pumped storage	16	2.7	1.00
OCGT	34	1.2	0.90
CCGT	124	26.7	0.90
TOTAL	272	74.4	

Table 3: Transmission connected conventional unit types [11].

The capacity outage probability table technique [2] assumes available capacity follows a Bernoulli distribution between zero and full capacity. With a 1 MW bin size, the resulting aggregate probability density functions have mean and standard deviation of 65.3 GW and 1.8 GW, respectively. Using this distribution the hourly winter LOLPs can be computed (3).

For simplicity each normalised hourly *d* is assumed to be fixed and does not itself follow an assumed probability distribution. Hourly LOLPs can be summed to produce the total 9.5-winter LOLE (4). Similarly the reduced LOLE is calculated using the expected wind output at each hour estimated by the numerator of (2).

3.3.2 Build-based capacity value: focus on offshore wind

Initially the offshore wind resource is considered in isolation with particular interest in the relationship between the spatial distribution of generation capacity and capacity value. The hourly aggregated GB offshore wind LFs are estimated using a projected offshore wind build schedule. This timetable is constructed from the three Crown Estate auctions (rounds 1-3) that define the locations and expected capacities of the offshore farms (see [13]). The aggregate LFs are then derived using the geographically weighted average of locational LFs (2). The results of this analysis are illustrated in Figure 5.

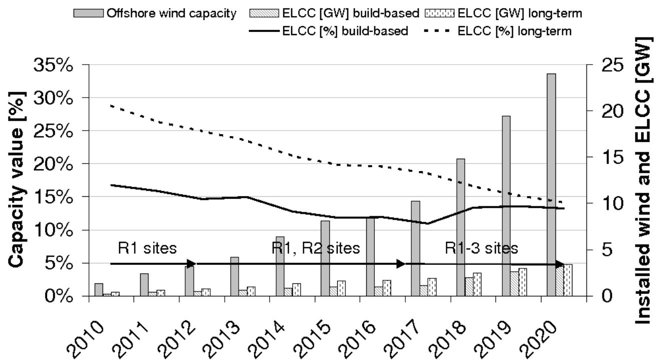


Figure 5. Capacity value (left y-axis) and ELCC (right y-axis) results for GB offshore wind using long-term and build-based LFs.

The dashed line shows capacity values calculated using *long-term* aggregate weighted LFs. This assumes all sites are included and the contribution from individual locations scaled by weighting their capacities. The solid line shows the capacity values calculated using the *build-based* aggregate weighted LFs over just the wind farm sites expected to be online at the start of each year considered. The total installed capacity expected to be online by the stated year is the same in both cases, however the build-based LFs are weighted across a less diverse resource. The graph demonstrates that considering sites by build schedule leads to lower estimated capacity values with the monotonically decreasing characteristic common in capacity value plots not present (e.g., Figure 7). This can be explained by the added value of capacity diversity improving (but not eliminating) the impact of dependence between sites on resource reliability in some years (particularly for the larger Round 3 sites).

Further, the box in Figure 6 shows the expected availability and standard deviation of offshore wind capacity during high demand hours (note that the total installed capacities are as in Figure 5). In this case, those demand hours within 5% of peak provide a sample size of 655 hours over the 9.5 winters. The selected probability mass functions demonstrate how the distribution of aggregate LFs for the sampled hours changes over time. There is a visible reduction in expected low LF hours as geographical diversity increases, and the probability of high loads factors remains high relative to those typically simulated for onshore (e.g., [1]).

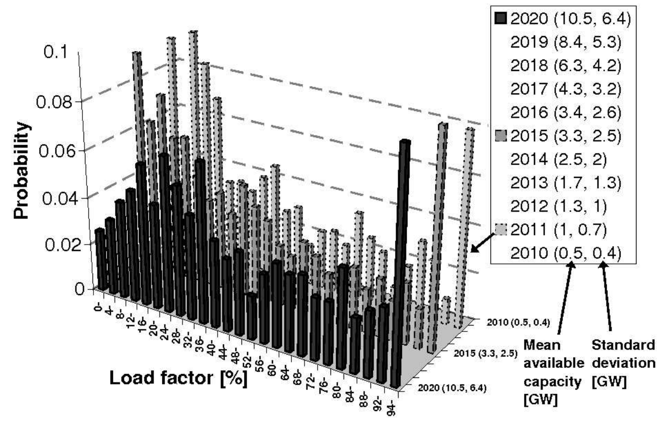


Figure 6. Probability mass function for GB off-shore wind LFs for selected years based on demand hours within 5% of annual peak. If the LF falls in a particular range (x-axis), it is deemed to be at the middle of that range (i.e., LFs in the range 0-4% are deemed to be 2%). Mean and standard deviation of total available capacity depicted in box.

3.3.3 Aggregate GB wind capacity value

Attention now turns to the combined GB wind resource, Figure 7 shows the ELCC and corresponding capacity value results for combined on- and offshore analysis using the long-term aggregate weighted LFs. These results suggest that for high levels of highly geographically diverse installed capacity, a capacity value of 10% appears credible.

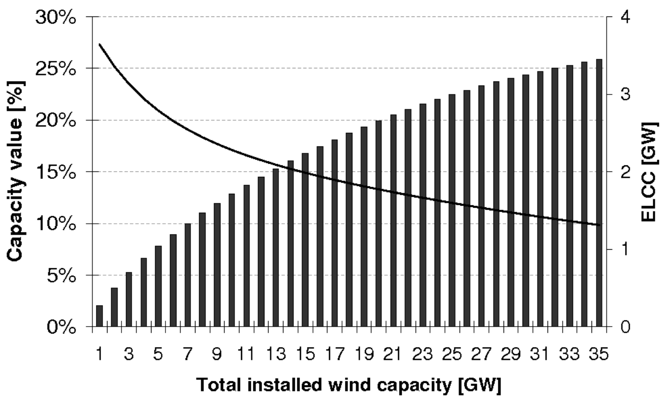


Figure 7. Capacity value (left y-axis, line) and ELCC (right y-axis, bars) for the combined on and offshore GB wind resource calculated with long-term aggregate LFs.

Figure 8 shows the results of a sensitivity analysis for ELCC that included the following (the numbers match Figure 8):

1) Scaling factor 0.69 applied to long-term offshore weighted loads factors;

2) Normalised peak demand reduced from 60 to 57 GW (initial LOLE reduced by 99%);
3) Reducing the total available conventional generation by 4 GW to a distribution with mean 62.2 GW, and standard deviation 1.7 GW (initial LOLE increases by 3000%).

As might be expected with sensitivity factor 1 the effect of scaling offshore wind LFs downwards tends to reduce the capacity value of wind. In 2 the risk is reduced, so is the ELCC and capacity value is also reduced. In 3 the risk increases and the ELCC does likewise. This is a well-known result of ELCC analysis and demonstrates the impact of underlying system risk on the results obtained.

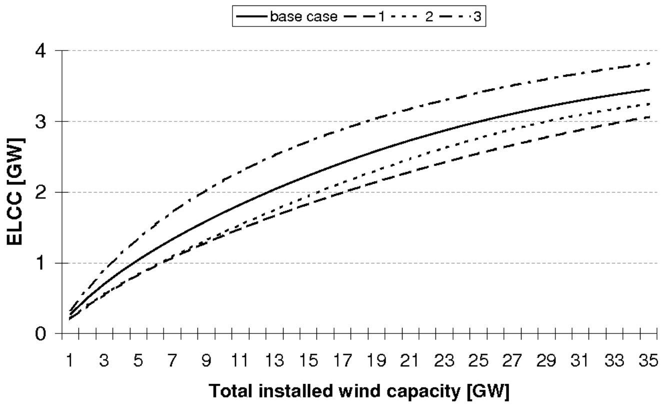


Figure 8. Sensitivity of ELCC value to assumptions.

4 Discussion

A major assumption here is that the last ten years winds are reasonably representative of future wind in the UK. Ten years is not enough to represent a full climatology and there is evidence of climate change affecting wind speeds [3]. However, ten years is long enough to sample a wide range of synoptic conditions and weather types and the analysis is a substantial improvement on comparable studies.

The output of wind at times of peak demand varies considerably between years. For example, in 2010, blocking high pressure over northern Europe led to very cold temperatures and high electrical demand, yet low wind speeds persisted over the UK. This highlights the difficulty, perhaps even the validity, of attempting to represent the contribution wind makes towards reliability as a single figure.

5 Conclusion

A detailed understanding of the wind resource and its variability in time and space is vital for understanding the contribution of offshore wind to reliable electricity supply. Here, a mesoscale atmospheric model was employed to create an ten year hindcast of British offshore wind speeds and simulated production.

Results of a reliability analysis have provided insight into the reliability of production from offshore wind at periods of high demand. What's more, credible estimates of combined long-term onshore and offshore capacity value have been derived and the sensitivities of these estimates to the underlying level of system risk discussed.

Acknowledgements

The authors acknowledge valuable input from Dr Chris Dent of Durham University. The PhD studentships awarded to S. Hawkins by the EPSRC Supergen Flexnet consortium and Kier Watson Trust and to D. Eager by the UK Energy Research Centre are gratefully acknowledged.

References

[1] P. Aguirre, C. Dent, G. Harrison, and J. Bialek, "Realistic Calculation of Wind Generation Capacity Credits," *CIGRE/IEEE PES Joint Symposium on Integration of Wide-Scale Renewable Resources Into the Power Delivery System*, July 2009.
[2] R. Billinton and R. N. Allan, *Reliability Evaluation of Power Systems. 2nd ed.* Plenum, 1994.
[3] G. P. Harrison, L. C. Cradden and J. P. Chick, "Preliminary assessment of climate change impacts on the UK onshore wind energy resource", *Energy Sources*, 30 (14), 1286-1299. 2008.
[4] Z. I. Janjic, "Nonsingular implementation of the Mellor-Yamanda level 2.5 scheme in the NCEP global model." *National Center for Environmental Prediction* Office Note, no 437. 2002
[5] Z. I. Janjic, "The surface layer in the NCEP Eta Model." In A. M. Soc (Ed.), *Eleventh Conference on Numerical Weather Prediction* (pp. 354-355). Norfolk, VA. 1996.
[6] A. Keane, M. Milligan, C. J. Dent, B. Hasche, C. D'Annunzio, K. Dragoon, H. Holttinen, N. Samaan, L. Soder, and M. O'Malley, "Capacity Value of Wind Power," *IEEE Transactions on Power Systems*, 26 (2), 2010, 564 – 572.
[7] J. B. Klemp, D. O. Gill, D. M. Barker, M.G. Duda, W. Wang, J. G. Powers, "A Description of the Advanced Research WRF Version 3". *National Center for Atmospheric Research*. Technical Note 2008.
[8] National Grid, "Average Cold Spell (ACS) Correction." nationalgrid.com/uk/sys_08/default.asp?action=mnch2_15.htm&sNode=2&Exp=N.
[9] National Grid, "Operational Data." nationalgrid.com/uk/Electricity/Data/Demand+Data/.
[10] National Grid, "Seven Year Statements," www.nationalgrid.com/uk/Electricity/SYS/.
[11] National Grid, "Winter Outlook Consultation 2010/11," www.nationalgrid.com/uk/Gas/TYS/outlook/.
[12] N. Screen, V. Parail, D. Sinclair, and O. Rix, "Electricity Market Reform Analysis of Policy Options," Redpoint Energy, Dec. 2010. www.decc.gov.uk/en/content/cms/consultations/emr/emr.aspx
[13] RenewableUK, formerly the British Wind Energy Association. www.bwea.com.
[14] S. Zachary and C.J. Dent, "Probability theory of capacity value of additional generation" www.supergen-networks.org.uk/filebyid/609/file.pdf. 2011

Validation of a dynamic control model to simulate investment cycles in electricity generating capacity

D. Eager, *Student Member, IEEE,* J. Bialek, *Senior Member, IEEE,* and T. Johnson

***Abstract*—The ability of the liberalised energy markets to trigger investment in the generation capacity required to maintain an acceptable level of security of supply risk has been - and will continue to be - a topic of much debate.**

Modelling the dynamics of investment in generation capacity can inform this debate. More precisely, if investment is viewed as a negative feedback control mechanism with energy prices acting as the feedback signal then the system can be formulated in terms of differential equations and addressed as a problem in optimal control.

The approach presented uses techniques from control theory to model investment market dynamics and a classical NPV approach is used for the investor decision process. The results of the model verification stage are presented whereby the model's ability to simulate the market trends witnessed in Britain since early 2001 is tested with encouraging findings reported.

***Index Terms*—Power generation economics, Optimal control, Simulation.**

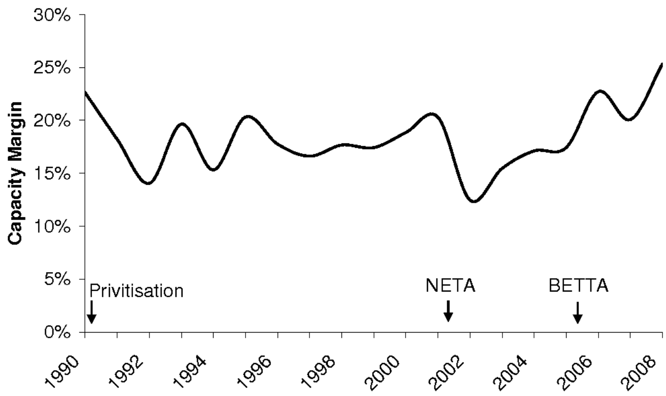


Fig. 1. Generation capacity margin in the GB since industry privatisation [7]. Dates of significant market framework changes also shown.

I. Introduction

So far, in Europe and the US, market liberalisation has been seen as a success in terms of efficiency and lower prices. However, various imperfections and catastrophes (e.g. Californian crisis in 2000 and 2001) indicate to market designers that improvements are still achievable. One of the main concerns is that the onus is now on privately owned generating firms to respond to supply shortages by investing in generation capacity in order to maintain an adequate level of security of supply risk.

An installed capacity capable of meeting demand whilst considering plant maintenance, unscheduled outages, utilisation factors, variable generation and unpredicted contingencies, is vital in order to avoid power shortages and blackouts. Nowadays it is widely accepted - certainly in Great Britain (GB) - that a benchmark capacity margin at or above 20% for a predomiately thermal system provides and acceptable level of risk. To illustrate this point, Fig. 1 shows the GB capacity margin in recent years; there has been a clear oscillation in the margin and what's more as the penetration of wind power increases, it will need to increase significantly if the risk level is to be maintained. Many observers attribute this capacity margin instability to the market framework under which the industry operates. Capacity payments (and such like) are used in most countries in order to ensure an adequate capacity margin, however GB is a rather rare (but not unique) example of an energy-only market, i.e. without capacity payments.

The purpose of this research is to establish whether a liberalised market framework is capable of inducing optimal investment in generation capacity over a long-term time frame (∼20-30 years). In the first instance, a simulation model of the current GB energy-only market has been constructed with the aim of estimating long-term characteristics and whether without capacity instruments it is likely to induce large price, reliability and capacity oscillations.

It is well known in control engineering that time lags and uncertainty are responsible for worsening system stability and creating undamped oscillations. Any generation investment has an inherent time lag of 3-5 years for gas and coal plants and 8-10 for nuclear before becoming operational. Moreover, analysis of consecutive GB System Operator (SO) Seven Year Statements [11] showed that many projects are subject to delay and/or abandonment. This combination of inherent investment time lags, delays and uncertainty about the level of capacity coming on line in the future creates generation capacity oscillations like those shown in Fig. 1.

By taking a dynamical systems approach, the simulation can be analysed for its stability characteristics and furthermore, controllers with the goal of minimizing the effect of uncertainties can be designed. Other models that share similarities with this methodology include [9, 15, 16, 20] and more significantly [14], which the work presented here is an extension of. Furthermore [9] and [20] explicitly state the dynamic equations used to model the evolution of the system, however crucially none go as far as to analyse the system for

Dan Eager is with the Institute for Energy Systems, The University of Edinburgh, EH9 3JL, UK, Phone: +44 (0)131 650 5596, email: d.eager@ed.ac.uk

Janusz Bialek is with the Energy Group, School of Engineering, The University of Durham, DH1 3LE, UK. email: Janusz.Bialek@durham.ac.uk

Tim Johnson is with the Department of Actuarial Mathematics and Statistics and the Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK. email: T.C.Johnson@hw.ac.uk

its stability characteristics.

Like many countries, this topic is currently very relevant in GB and a large volume of work has been published in this field by both academia and industry [1, 2, 19, 21]. These studies focus on the capability of the current market framework to deliver a secure electricity mix whilst meeting firm emissions reduction targets and renewable energy obligations. In the case of [19], a static approach to modelling is taken whereby an investment forecast is made assuming future gas and coal prices and then key indicators such as electricity prices are calculated. However in the model presented below a feedback mechanism is present, whereby a year-on-year investment decision is taken enabling investors to adapt their behaviour based on current market conditions; this includes projects in the pipeline, time delays, fuel prices and so on.

The results presented here are from the model verification stage of the work; a comparison between the market dynamics witnessed since the introduction of the New Electricity Trading Arrangements (NETA) in England and Wales (E&W) and the simulation results over the same time period are presented. NETA is the name given to the system under which electricity was traded in E&W up until April 2005. It came into being in March 2001 and was altered in April 2005 when the electricity market in E&W merged with its counterpart in Scotland into a single GB wholesale market[1].

The paper is organised as follows: Section II describes the model formulation and the important elements of the simulation; this includes how capacity, demand, fuel price and wholesale electricity price are simulated. Section III contains a description of the investment decision process and how uncertainties faced such as fuel price and demand are modelled. Section IV includes the results of the comparison study and finally section V concludes the paper.

II. The Model

The approach of applying techniques from optimal control theory is taken to model market investment dynamics (Fig. 2). Because the model is dynamic, current prices and their predictions are fed back to the investment block modifying the investment behaviour. The resulting investment decisions are then fed back to the pricing mechanism hence closing the loop.

Two control problem models have been developed in parallel for this work. One is the discrete-time model formulated with difference equations and the other is the continuous-time model formulated with differential equations.

A daily time-step model has been implemented for both the discrete and continuous case. The model using the daily time-step is run to establish likely investment trends for the years ahead. The results from this simulation are added to the hourly time-step model and a detailed analysis of key factors such as wholesale price trends and security of supply risk are assessed. In fact a direct comparison between historical wholesale prices and simulated prices is included in order to verify the model

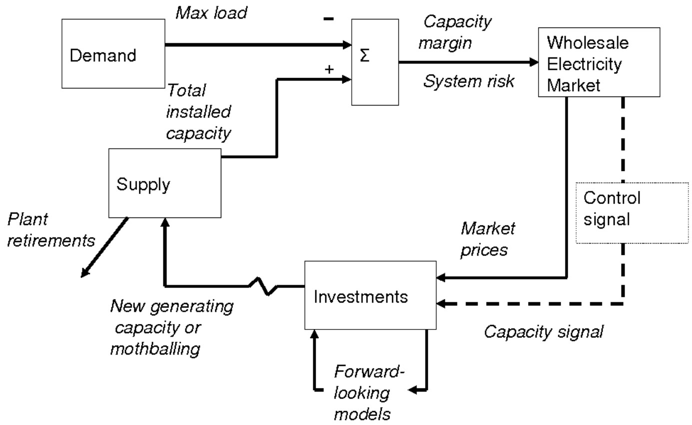


Fig. 2. Electricity investment as a control problem; investment can be viewed as a negative feedback control mechanism with energy prices (as a function of generation capacity margin) acting as a feedback signal.

pricing function formulation, details on this function are given in section II-C.

Because there is an investment element to the model, a forward-looking simulation has also been developed. This model uses market conditions in the real-time simulation as initial conditions and then makes predictions on the future state of the system during the lifetime of a potential investment. Crucial uncertainties such as future demand, fuel prices and wholesale energy prices are all simulated.

Therefore, for the purposes of model verification, the real-time model is based on historic data with minimal uncertainty (except for new investments and price), however the investment decisions are still based on a uncertain perception of the future.

A. Capacity

The central dynamics of the system surround the evolution of installed generation capacity. The rate of change in capacity at a particular time-step is dependent on new plant coming online or being demothballed together with any plant retiring or being mothballed. Both are delayed signals from some earlier time. In the case of new plant this delay is build time, and in the case of retiring plant this delay is lifetime. The control system linear time-invariant (LTI) differential equations are given by

$$\frac{dG_x}{dt} = \frac{dI_x}{dt} - \frac{dR_x}{dt}, \tag{1}$$

where G_x is the installed capacity of plant type x,

$$\frac{dI_x}{dt} = I_x(t - \tau_x + \epsilon) = \begin{cases} C_x & if\ invested\ in\ x \\ 0 & otherwise \end{cases} \tag{2}$$

is the investment rate and

$$\frac{dR_x}{dt} = I_x(t - \alpha_x) = \begin{cases} C_x & if\ x\ retired \\ 0 & otherwise \end{cases} \tag{3}$$

is the rate of plant retirement. C_x is the capacity, τ_x is the expected build time, α_x is the expected lifetime and ϵ is a random variable representing unforeseen delays (e.g. in the planning process). An aggregate approach is taken whereby

[1]Namely the British Electricity Trading and Transmission Arrangements (BETTA).

capacity is combined into five technology tranches, namely nuclear, coal, wind CCGT and OCGT each with its own cost characteristics.

In the base case, project delays are modelled as lognormally distributed with mean 1 year and variance six months. Furthermore, the number of projects that are abandoned is modelled using a simple discrete probability distribution such that on average $x\%$ of projects are abandoned. In the base case, $x = 0$ for coal nuclear, CCGT and OCGT[2] and owing to its poor record at gaining planning permission and so on, $x = 50$ for wind[3].

Finally, plant included in the Large Combustion Plant Directive (LCPD)[4] is also modelled whereby the lifetime of the 11GW+ which has opted out of the directive (i.e. opted to close rather than retrofit the equipment necessary to reduce emissions) is given a reduced lifetime based on the estimations of remaining generating hours given in [13].

B. Fuel Prices

The data shown in Fig. 3 was used to model generator fuel and emissions prices at real-time. Plainly the gas price is modelled with a greater degree of accuracy than the other fuel prices, this was for a number of reasons: firstly only detailed historical data of gas prices was obtainable, secondly by having a detailed model for gas, the characteristic of electricity prices being strongly correlated with gas prices can also be checked and verified.

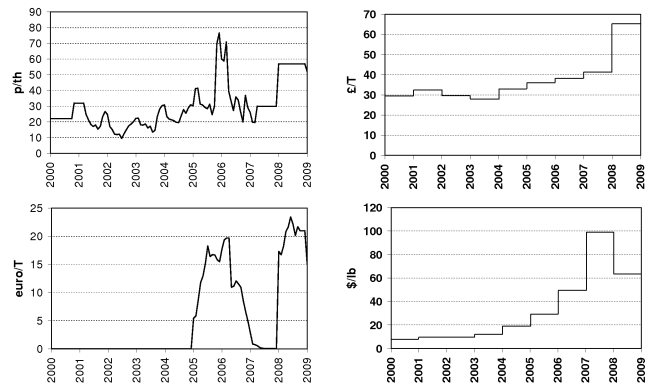


Fig. 3. Fuel prices, from top left (clockwise) gas [22], coal [3], uranium [8] and carbon [22].

C. Wholesale electricity price simulation

Simulating wholesale energy prices is an important aspect of any model that addresses investment in generating capacity under a liberalised market framework. Furthermore, the wholesale price trends witnessed can indicate to observers how well the system is functioning and whether there is has been adequate investment in capacity. For example, there is a tendency for over-investment to lead to periods of low energy prices and conversely under-investment leads to periods of high prices.

[2]The logic here being that traditionally, new thermal capacity is usually built on an exisiting site with a grid connection so will not come up against stiff planning resistence.

[3]At the time of writing this paper, a basic analysis of British Wind Energy Association (BWEA) data [5] revealed that from 2004 to 2008, 495 applications were submitted and 258 have been approved to date.

[4]A control on emissions from heavily polluting large combustion plant introduced by the EU in 2001.

TABLE I
GENERATOR DATA [17].

Gen. type	Efficiency	Availability	Capital £/kW	O&M fixed £/kW	O&M variable £/GJ	life-time yrs	build time yrs
Nuclear	0.36	0.78	1485	43	0.05	40	7
Coal	0.48	0.86	400	14	0.28	40	5
Wind	-	0.28	1000	27	0.0	25	4
CCGT	0.58	0.87	300	7	0.56	25	3
OCGT	0.39	0.95	330	10	0.75	40	2

It is assumed that in a competitive energy market, generators will bid to produce electricity at or around their marginal cost and therefore our pricing function has a marginal cost element included. Also during periods of tight supply, an upturn in prices is commonplace, therefore an uplift function which is driven by the generation capacity margin is also included. Details of each element are described below.

1) The marginal cost element: This is computed using the generator supply curve. Individual supply functions are derived from the generator marginal costs; in general these are based on the common cubic function of power output [24]. By assuming a uniform heat rate across the operating range, the cubic model is simplified and only the linear term remains:

$$C(P) = bP \cdot F + V \cdot P, \tag{4}$$

where P is the power produced (MW), F is the fuel cost, V is the variable operation and maintenance cost and b can be derived from the generator thermal efficiency.

The cost of carbon is included by appending an additional term to the cost function (4) of the form $CO = bP \cdot \vartheta \cdot F_{car}$, where ϑ_x is the carbon produced by burning the particular fuel type and F_{car} is the price of carbon. The full generator cost function is then given by

$$C(P) = bP \cdot (F + \vartheta \cdot F_{car}) + V \cdot P. \tag{5}$$

The assumptions for each of the plant types outlined are given in table I. The CO_2 equivalent output of each generator type is given by: gas - 185 (Kg/MWh) and coal - 330 (Kg/MWh) [6]. All other technology types are modelled with zero carbon emissions.

2) The uplift function: Two forms of uplift function were considered; an exponential [23]:

$$U_1(t) = e^{a \cdot L(t) - C(t)} \tag{6}$$

and a hyperbolic [14]:

$$U_2(t) = \frac{s \cdot g \cdot L(t)}{C(t) - g \cdot L(t)} \tag{7}$$

where s, b and l are scalar factors and can be set in order to induce the desired price behaviour and $L(t)$ and $C(t)$ are the load and available capacity respectively (in GWs).

The hyperbolic function was used under reasonably healthy system conditions when an adequate capacity margin is present

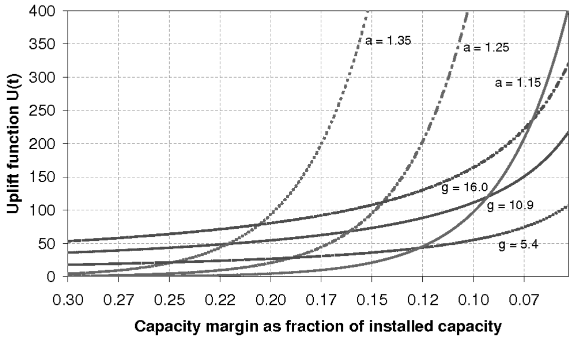


Fig. 4. Example electricity uplift function simulations for both (6) (labelled with a parameter) and (7) (labelled with g parameter).

(range 50-20%) and the risk is low. However it does not increase sufficiently quickly as the system approaches scarcity. Under these conditions, the exponential function is better. When there is inadequate generation, the system is at the mercy of suppliers who can bid any price they can think of. This is referred to as the *hockey stick* bid; a normal supply curve is bid for most of the generation but the last few MW are bid at a very high price. An exponential function models this situation quite well. Therefore the uplift function was constructed as a max of (6) and (7):

$$U(t) = max\left\{U_1(t), U_2(t)\right\} \tag{8}$$

3) Price setting procedure: The price is set by performing an economic dispatch and the marginal cost bid (5) of the last generator type to be dispatched sets the marginal cost element of the price. By assuming a linear cost curve, the speed of the economic dispatch optimization model is greatly increased; the supply curve is simply a piecewise step function where each piecewise constant segment, S_x, is defined by

$$S_x = \frac{dC}{dP} = b \cdot (F + \vartheta \cdot F_{car}) + V. \tag{9}$$

for each generator type x.

The current demand and available capacity is then fed into (8) and the wholesale market price is given by:

$$S(t) = C_x^{marg}(P) + U(t) \tag{10}$$

where $C_x^{marg}(P)$ is the marginal cost of the final generator type to be dispatched.

III. Investment Decision

In order for adequate capacity to be maintained, investment in new capacity must be forthcoming. In this model the investment decision is taken annually and is based on the Net Present Value (NPV) of expected future profits. These profits are calculated in the standard way: $\Omega = \pi \cdot P - C(P)$, that is profit ($\Omega$) is revenue received from selling power ($\pi \cdot P$) minus the costs incurred producing it ($C(P)$).

To give significant weight to the early years of an investment, only the revenues for the first 15 years of a project are considered, however expected costs incurred are included for the lifetime of the plant. Renewable obligation certificates[5] (ROCs) becomes active at the end of 2002 and nuclear is not considered for investment before 2006[6].

A forward-looking simulation model is used to assess the future state of the system. This model is formulated in the same manner as the real-time simulation (1). To address the problem of future uncertainties and imperfect investor foresight, a Monte Carlo approach is taken when calculating an expectation of the future.

When calculating the yearly utilisation factor of a potential investment based on future load duration curves (LDCs) and plant availability, a 5^{th} order polynomial fit is applied to the LDC and the required baseload (assumed to be value of the LDC at max duration) and peak energy is computed. The forward-looking model assumes an annual growth rate of the base demand profile (based on 2001 data) of 1.2% with forecast volatilty of 100MW increasing at a rate of 10% p.a. This volatility is modelled as a random variable $\sim N(0, 10^2 \cdot (1.1)^{y-1})$ where y is the simulation year.

The expected price of each of the fuel types in the system is an exogenous parameter and is modelled as a stochastic process. More precisely a Geometric Brownian motion

$$dF_t = \mu(F_t)dt + \sigma(F_t)dW_t, \quad F_0 = f > 0. \tag{11}$$

is constructed for each fuel type and the parameters are defined based on perceived drift (average behaviour) and volatility characteristics. The initial values ($F_{x,0}$) are taken to be the average of prices witnessed over the previous year. Forward curves are not included in this iteration of the model.

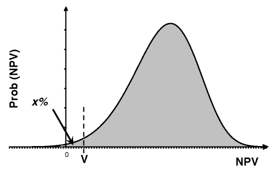


Fig. 5. Example of NPV confidence test showing optimal V (i.e. V_x^{opt})

For each Monte Carlo simulation, a different NPV is computed, say V_x, and a $V_x > V_x^{opt}$ check is carried out based on a statistical confidence test (Fig. 5), where is V_x^{opt} is the minimal acceptable V_x for the particular technology. There is also a secondary logical check that the expected utilisation factor for all thermal generation is greater than zero in the first year of operation ($uf(1) > 0$), otherwise the investment is delayed. When a number of technology types are optimal in terms of the above criteria, those projects are ranked by their Profitability Index (PI) and the option with the largest PI is chosen. PI is defined as the ratio of present value (PV) of cash flow and PV of initial investment.

We do not consider different types of investor being present in the market at this stage. Instead we model a single investor who is risk averse; this investor assumes all plant currently in planning or under construction will come online with 100% certainty and furthermore the minimum NPV requirement

[5]UK mechanism to stimulate investment in renewable generation whereby each MWh generated from a renewable source is eligible for subsidy.

[6]This reflects the UK's political stance on nuclear held at the time.

is set at twice the project fixed costs and accepted if the confidence test on V_x is significant at the 95% level.

In the case of peakers, revenue from the SO annual tender for short-term operating reserve (STOR) is also considered. Based on the data at [12], a 2GW annual tender with an availability price of 2 £/MW/h with a utilisation (estimated at 3% of hours) price of 100 £/MWh was included.

The model also enables peaker plant to be mothballed. The decision to mothball plant is taken every six months and is based on $V_x < V_x^{opt}$ where V_x^{opt} is the generator fixed costs for next six months. The mothballing of existing capacity is a direct consequence of the market feedback mechanism; if a genco believes that a particular plant will not be able to cover it's fixed costs from revenues in the energy-only market (and assuming it has not established a contract with the SO to provide reserve), then mothballing the plant until the price rises sufficiently is an option.

Once capacity is mothballed (of which there is currently 1.25GW in GB [13]) then it remains connected to the system but no longer contributes to short-term security of supply risk calculations (as it would take a number of months to get the plant ready to generate again from the mothballed state).

IV. Simulation Results

The model has been implemented using in the Matlab/Simulink environment.

To verify our approach and model formulation, we have attempted to simulate market dynamics in GB since the introduction of NETA. The model has been tuned to reflect the situation just prior to the introduction of this market framework. Real-time demand is modelled based on historic half-hourly GB demand data taken from [10]. The initial plant mix used the simulation is shown in table II, also shown are projects already under construction at the start of the simulation[7]. To get a complete GB picture, we have combined the E&W and Scotland systems into a single energy market[8]. As mentioned previously, an LTI equation is required for each generator type in the model (1), therefore the model simplifies the actual GB system by only considering five generation technologies. In order to achieve the same installed GB capacity as in reality, capacity from peaker plant such as pump storage and oil are combined with OCGT and other baseload such as natural flow hydro is included with coal. The dual fuel plants are divided based on their estimated thermal efficiency characteristics and load following capabilities [4]. By taking this approach, model complexity is kept to a minimum and computational time is reduced. For example, using an Intel dual-core 2.40GHz processor with 3.12GB RAM with 100 Monte Carlo simulations for each investment decision and a daily time-step simulation period of 8 years takes many hours to complete.

[7] That is projects whose time to completion was below the base build times given in table I.

[8] As mentioned earlier, this actually became the case in April 2005 with the introduction of the British Electricity Trading and Transmission Arrangements (BETTA) in GB.

TABLE II
Approximate GB installed capacity in 2000 [7].

Plant type	Capacity (GW)	Under construction	In the model
Nuclear	12.49		12.5
Coal	24.84		30.5
Hydro	1.33		
Mixed or Dual Fuel	6.87		
CCGT	19.35	1.27	20.6
Wind	0.12^a	0.8	0.1
OCGT	1.32		8.35
Oil	2.94		
PS	2.79		
TOTAL	72.04	2.07	72.05

[a] transmission connected capacity only

A. Base case results

Fig. 6 shows the model simulation of capacity margin against reality. Plainly the negative response after 2001 of the solid line is deeper but the frequency of oscillation in the two lines is very similar and the margin also seems to be better damped than in reality; however considering the model is based on a number of assumptions and simplifications, these results show a good agreement with reality.

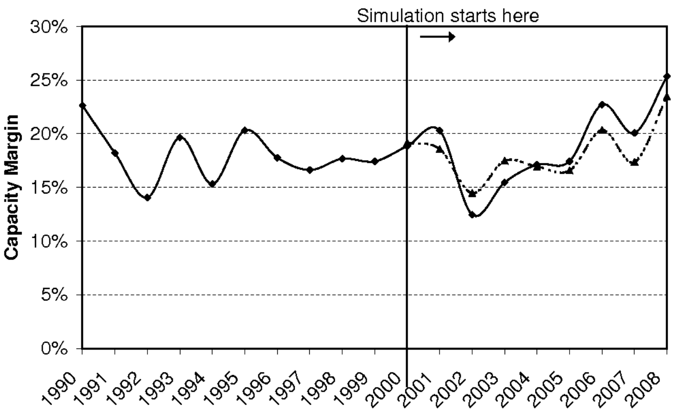


Fig. 6. Generation capacity margin oscillations witnessed since market liberalisation (solid line) and simulation results from 2000 (dashed line).

Fig. 7. Evolution of total installed capacity in GB (reality) and the simulation. GB maximum demand also shown.

Fig. 7 shows the evolution of installed capacity (ICAP) in the simulation versus reality in the base case (where the

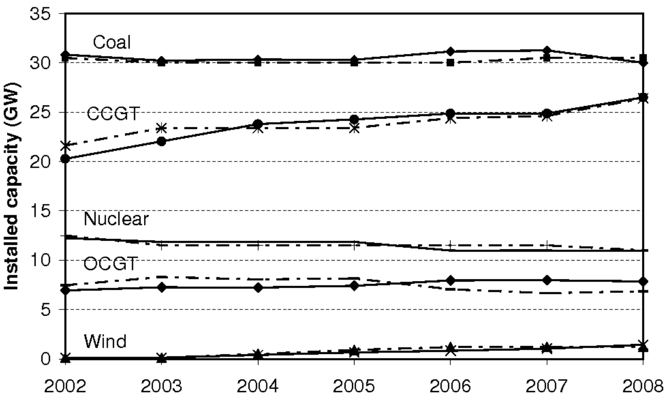


Fig. 8. Evolution of installed capacity for each generator type in GB (solid lines) and in the simulation (dotted lines).

investment decision in annual). Both lines follow a similar trajectory however the simulation is much smoother, this effect is created by plotting the installed capacity at beginning of each year. As mentioned already, prices witnessed in the market are fed back into the investment decision, however by having an annual investment decision time-step only averages are considered.

B. Investments

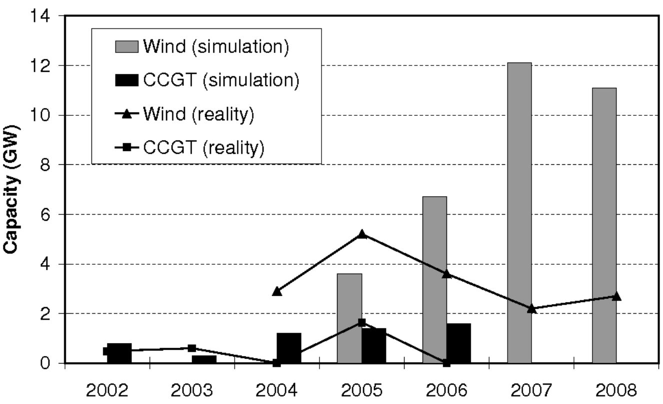


Fig. 9. Capacity investment triggered during the simulation and actual investments announced in GB. Data on wind from BWEA [5] and CCGT builds estimated from SO seven year statements [11] assuming 3 year build time.

Fig. 9 shows the volume and type of investment triggered during the simulation. There is little investment in the early years owing to the 2GW already being under construction when the simulation starts and furthermore investors would not be looking to undertake a new investment when prices are falling and uncertainty increasing.

Wind power investment is also another key area of interest. Fig. 9 shows that no significant investment in wind occurs until after 2005. This is owing to not only the ROC subsidy becoming active, but also there are significant extra costs incurred for other projects; for CCGT the cost of gas increases by 60% from 24 p/therm to 39 p/therm and coal projects have the additional cost of emmissions to consider. Fig. 9 also shows BWEA data on applications received for new wind farms received in GB [5], the main points to note here are although wind power investment is delayed until 2005 - perhaps owing to the high hurdle rate for this technology, there seems to be an over investment in later years. Recalling that the model does not consider grid connection and so on - so conceptually this system is single bus - arguably this trend is quite plausible. Of course much of this investment will not make it through the build stage owing to the high probability (0.5) of project abandonment.

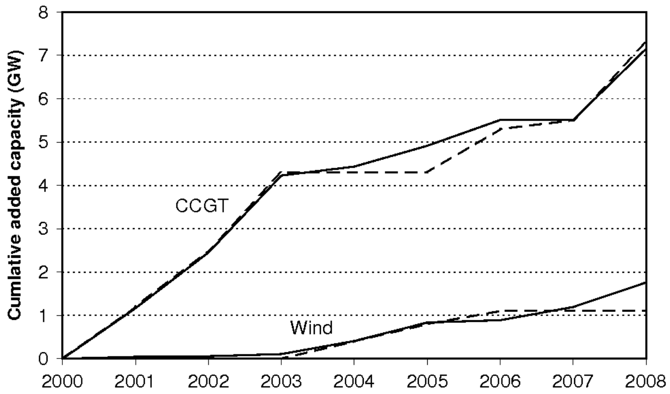


Fig. 10. Cumulative capacity added to the system in the simulation (dotted-line) and in GB (solid line).

The cumulative capacity added to the system can be see in Fig. 10. New builds were initialised for both wind and CCGT and both represent reality well (of course some of these builds were already underway at the simulation start time, but those projects that were initialised by the investment decision process do contribute in the later years).

C. Mothballing

Although in reality CCGT capacity growth remains constant throughout the simulation (Fig. 10), the same cannot be said of OCGT plant; inspection of Fig. 8 shows a downturn in OCGT capacity in 2005 not akin with reality. This downturn can be explained by a sharp increase in mothballed plant in 2005 and thus the installed capacity also falls behind reality (Fig. 7). The likely cause of this behaviour stems from a combination of an increase in installed CCGT capacity - which reduce OCGT revenues from the energy market, together with a 60% increase in gas prices (cf. Fig. 3). Limited information is available pertaining to when exactly GB's (currently) 1.25GW [13] of plant was mothballed. The information that was available concerned the mothballing of plant in early 2003, with much of it returning to the system in the winter of the same year [18]. Owing to the reduced complexity of the model compared with reality, it is difficult to capture this behaviour as individual plant characteristics cannot be included. Therefore although the model is replicating the dynamics of the system quite well, some further work is required in the area of mothballing.

D. Importance of the reserve market

A sensitivity test was carried out whereby a weighting coefficient, w, in the range 0 to 1 was applied to the additional revenue received by OCGT plant for providing STOR. For example, $w = 1$ is the base case and $w = 0$ represents a complete removal of the STOR market. Fig. 11 shows the importance of having a separate market for reserve (which is essentially a capacity payment for peaking plant); without it the initial capacity margin slump is deeper owing to intensified withdrawal of peaking plant from the system. Owing to greater CCGT investment in the early years than in the base case, the margin does recover after 2005, however CCGTs are not well suited to providing reserve and hence the level of security of supply risk is greatly increased. Interestingly the introduction of w leads to a better initial downturn in capacity at the start of the simulation (w=0.5) and although the margin remains below reality in subsequent years, the difference is constant implying that this payment is a key parameter in the model. On the whole, including a steady revenue stream for peakers removes much of the uncertainty surrounding utilisation and price, thus keeping more peaker plant connected to the system.

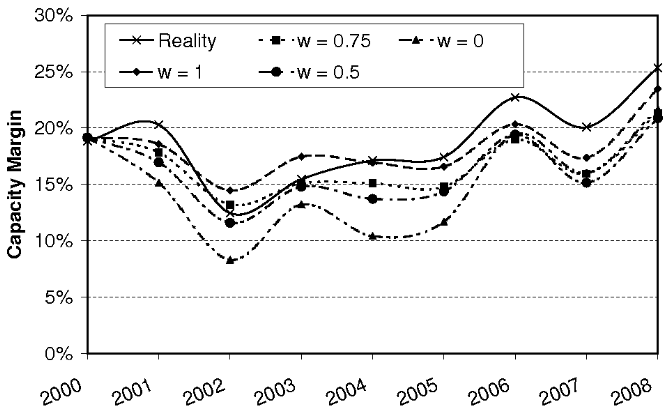


Fig. 11. Plot of GB generation capacity margin (solid line) with reserve market sensitivities (dashed lines) for various coefficient weightings (*w*).

E. Wholesale energy prices

Another key area of analysis was the ability of the model to simulate wholesale energy prices. Fig. 12 shows a good overall matching of trend. The loss of precision after April 2008 can be explained by recalling that detailed gas price data for these years was not available (cf. Fig. 3). The ability to push through high fuel costs is demonstrated when the simulated prices for April 2001 to April 2007 are viewed in isolation. The mean difference between the actual and simulated prices was 4.0 £/MWh with standard deviation 9.6 £/MWh, however the reduced dataset with more accurate fuel price models was 0.70 £/MWh with standard deviation 4.35 £/MWh.

It should be emphasised that the main aim of the model is not to predict the electricity prices but to investigate whether or not an energy-only market results in excessive capacity oscillations and if so, design an appropriate damping mechanism. Hence our main aim is to investigate the need for a capacity mechanism whatever happens to the primary fuel prices. Obviously as the absolute level of electricity prices depends strongly on the primary fuel prices, predicting future electricity prices would require a sophisticated gas and coal price prediction model and that would be beyond the scope of this research. Any changes in primary fuel prices have the effect of shifting the level of prices up or down (assuming constant spark and dark spread) and could cause a switch from gas to coal or vice versa. This is justified by inspection of Fig. 13 and 14 which show results from a simulation using predicted gas and coal prices and not actual data from 2001-2007. Plainly the price graph is different (influence of demand profile is more noticeable) however the capacity graph is only marginally affected.

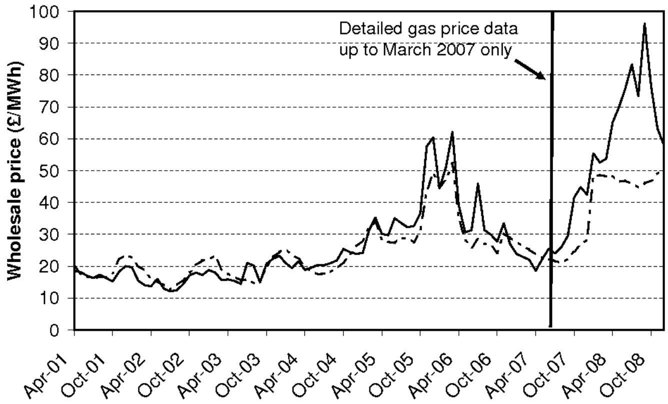


Fig. 12. Average monthly wholesale energy prices in the simulation (dashed line) and in GB wholesale market (solid line) [22].

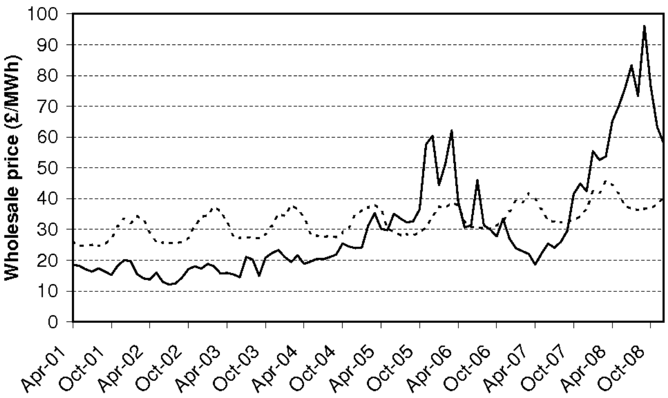


Fig. 13. Average monthly wholesale energy prices in the simulation when using fuel price predictions (dashed line). GB wholesale market data in included for a comparison with Fig. 12 (solid line) .

V. Conclusion

A power generation investment model which is formulated as a problem in optimal control has been presented. As when analysing any control system, the first task is to model the behaviour of the real system and simulations against the market dynamics witnessed in GB since the introduction of NETA in 2001 are very encouraging. Furthermore, the investment decision aspect of the work is seemingly rational in it's decision making process, which bodes well for the next stage of the work - that is to analyse the system for it's stability characteristics and move on to looking at expected future investment dynamics in GB. The next iteration of the

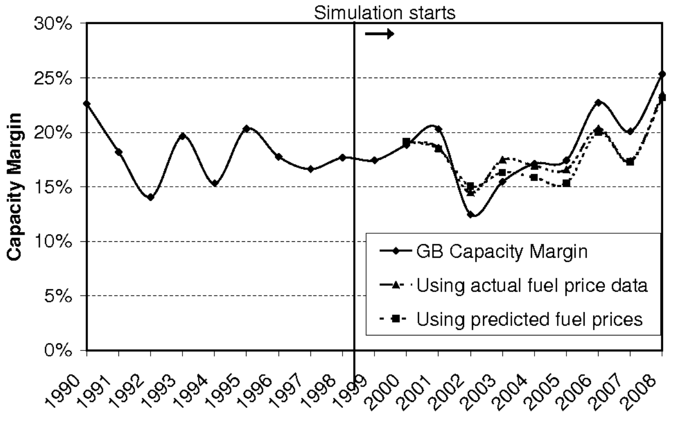


Fig. 14. Generation capacity margin oscillations witnessed since market liberalisation (solid line) and simulation results from 2000 with actual prices (dashed line, as in Fig. 6) and predicted prices (dotted line).

model will also aim to include different types of investors and assess how alternate perception of reality influence the model feedback. From here one can identify problems such as instability and look to design and implement an economic controller with the goal of meeting system performance specifications; in this instance these specifications surround the system level of security of supply risk.

ACKNOWLEDGMENT

The authors acknowledge valuable discussions with C. Dent, C.M. Gibson, G.P. Harrison and the National Grid Company.

REFERENCES

[1] Implementation of eu 2020 renewable target in the uk electricity sector: Renewable support schemes. Technical report, Redpoint Energy, June 2008.

[2] BERR & Ofgem. *Energy Markets Outlook*. 2007.

[3] BERR, Quarterly Energy Prices, March 07 and Sept. 09, http://www.apxgroup.com.

[4] BM unit data, BM reporting, http://www.bmreports.com/.

[5] BWEA, http://www.bwea.com/ukwed/index.asp.

[6] Carbon Trust, http://www.carbontrust.co.uk/resource/conversion_factors/default.htm.

[7] Economic and Social Data Service, https://www.esds.ac.uk/international/access/dataset_overview.asp#desc_ieaei.

[8] First Uranium, http://www.firsturanium.com/.

[9] Gary, S., Larsen, E. R. Improving firm performance in out-of-equilibrium, deregulated markets using feedback simulation models. In *Energy Policy*, volume 28, pages 845–855, 2000.

[10] National Grid. Operational data. http://www.nationalgrid.com/uk/Electricity/Data/Demand+Data/.

[11] National Grid. Seven year statements. http://www.nationalgrid.com/uk/Electricity/SYS/.

[12] National Grid. Stor. http://www.nationalgrid.com/uk/Electricity/Balancing/services/reserveservices/STOR/.

[13] National Grid. Winter outlook reports. http://www.nationalgrid.com/uk/Gas/TYS/outlook/.

[14] Häni, T., Bialek, J.W., Cherkaoui, R. Modelling generation capacity margins in market environments as a dynamic control problem. In *IEEE Power Engineering Society General Meeting*, 2006.

[15] Shen He, Lin Xiao, Jing Zhang, Guoyue Liu, and Ming Zeng. Capacity mechanism analysis based on dynamic simulation of generation investment. In *IEEE/SOLI 2008. IEEE International Conference on*, volume 1, pages 265–269, Oct. 2008.

[16] Kayoda, T., Ihara, S., Larose, E., Sanford, M., Graham, A. K., Stephens, C. A., Eubanks, C. K. Utilizing system dynamics modeling to examine impact of deregulation on generation capacity growth. In *Proceedings of the IEEE*, volume 93, Nov 2005.

[17] UK MARKAL 2007 Electricity Generating Technologies, http://ukedc.rl.ac.uk/data.html.

[18] Ofgem. Security of supply october 2003 to march 2004, September 2004. www.ofgem.gov.uk.

[19] Ofgem. Project discovery: Energy market scenarios, October 2009. www.ofgem.gov.uk.

[20] Olsina, F., Garcés, F., Haubrich, H.-J. Modeling long-term dynamics of electricity markets. In *Energy Policy*, volume 34, pages 1411–1433, 2006.

[21] UK Energy Research Centre. *Investment in Electricity Generation*, chapter 2-3. May 2007.

[22] APX, http://www.apxgroup.com.

[23] Visudhiphan, P. Dynamic investment in electricity markets and its impact on system reliability. In *Market Design 2001 Conference, Stockholm*, pages 91–110, 2001.

[24] Wood, A.J., Wollenberg, B.F. *Power Generation, Operation and Control*. Wiley, 1996.

Dan Eager Student Member M '08. Was born in Brighton, UK, in 1982. He received the B.Sc. degree in mathematics and computer science from The University of Sussex, Brighton, in 2004, and the M.Sc. degree in operational research from The University of Edinburgh in 2007. He is currently undertaking a Ph.D. within The University of Edinburgh's Institute for Energy Systems under the supervision of Janusz Bialek and Tim Johnson.

Janusz W. Bialek Is Professor of Electrical Power and Control at Durham University, UK. He holds M.Eng (1977) and Ph.D. (1981) degrees from Warsaw University of Technology, Poland. In 1981-1989 he was with Warsaw University of Technology, in 1989-2002 with Durham University, UK, and in 2003-2009 with The University of Edinburgh. He has co-authored 2 books and over 100 technical papers. His research interests include power system economics, power system dynamics and sustainable energy systems.