Topics in deregulated electricity markets

Zili Li

Master of Economics and Finance

Submitted in fulfilment of the requirement for the degree of

Doctor of Philosophy

School of Economics and Finance

Queensland University of Technology

2016

Abstract

The successful operation of the electricity market depends crucially on understanding the complex interactions between demand, supply and infrastructure as well as the strategic behaviour of market participants. Achieving such an understanding is a daunting task which is made even more demanding in the presence of market deregulation. In seeking to contribute to a deeper understanding of the National Electricity Market (NEM) in Australia, and particularly the Queensland region, this thesis explores three major issues of considerable importance to both the market operator and to market participants, namely, load forecasting, the interaction between transmission infrastructure and price, and strategic bidding behaviour by generators. Turning first to the issue of load forecasting, the approach adopted in this thesis is to build a multiple-equation time series model for the purposes of forecasting load, paying particular attention to the way in which the seasonal structure present in load can be used to enhance day-ahead point forecasts. While point forecasts are useful to market operators, additional information about load uncertainty can be obtained by forecasting load quantiles. Using a similar model specification to that used for generating point-forecasts, two quantile regression models of load are formulated, estimated and used for forecasting quantiles. Moving beyond the demand side, the behaviour of wholesale prices is investigated with special attention paid to its connection with one of the defining physical characteristics of electricity market, transmission infrastructure. The effects of transmission constraints are shown to be an important determinant of wholesale price behaviour, particularly in the upper quantiles of the price distribution. Finally, the bidding behaviours of generators in response to extreme price movements are explored. In particular, the existence of extreme price events is found to promote strategic bidding and rebidding on the part of generators, practices which

are shown to be injurious to the operation of the market. Some policy options for dealing with these issues are proposed.

Keywords: Load forecasting; electricity price; transmission constraints; deregulation; strategic bidding and rebidding.

Statement of Originality

The material in this thesis has not previously been submitted for qualification at any other institution. To the best of my knowledge, this thesis contains no material previously published or written by another person except where due reference is made.

QUT Verified Signature

Zili Li, September 2016

Acknowledgements

I would like to express my gratitude to my supervisors, Professor Adam Clements and Professor Stan Hurn, for cultivating my research interests in electricity markets and statistical modelling.

Table of contents

List of figures **xi**

List of tables **xix**

1 Introduction **1**

1.1 Motivation 1

1.2 The structure of the NEM 2

1.3 The research questions 6

1.4 The structure of the thesis 8

1.5 Key contributions 10

2 The relevant literature **15**

2.1 Load forecasting 15

2.2 Price behaviour 19

2.3 Strategic bidding and rebidding 23

2.4 Conclusion 27

3 Forecasting day-ahead electricity load using a multiple equation time series approach **29**

3.1 Introduction 29

3.2 A prototype multiple equation model 31

3.2.1 The basic model structure 32

3.2.2 Dealing with the effect of temperature 33

3.2.3 Estimating and forecasting the prototype model 35
3.3 Extensions to the prototype model 37
3.3.1 Addressing seasonality 37
3.3.2 Intra-day correlations 42
3.4 A comparison of approaches to modelling seasonality 45
3.5 Assessing forecast performance of the full model 48
3.6 Conclusion 51

4 Forecasting quantiles of day-ahead electricity loads 53
4.1 Introduction 53
4.2 Specification for the quantiles 55
4.3 Quantile estimation with a parametric residual distribution 58
4.3.1 Estimation procedure 59
4.4 An alternative model with a non-parametric residual distribution 62
4.5 Forecasting and performance evaluation 65
4.6 Conclusion 71

5 The effect of transmission constraints on electricity prices 75
5.1 Introduction 75
5.2 Institutional background 77
5.3 Preliminary analysis 83
5.3.1 Explanatory variables 83
5.3.2 Regression models 86
5.4 Modelling the quantiles 90
5.5 Conclusion 97

6 Strategic bidding and rebidding in electricity markets 99
6.1 Introduction 99
6.2 Institutional framework 101
6.3 The anatomy of price spikes 104

6.4 Strategic bidding 108
6.5 Rebidding 115
6.6 Policy implications 121
6.7 Conclusion 124

7 Concluding remarks **127**

References **131**

List of figures

2.1 Averaged half-hourly load over a day and averaged half-hourly load over a week in Panels (a) and (b) respectively, for Queensland over the period from 12th July 1999 to 27th November 2013. 16

2.2 Queensland load and temperatures, from July 1999 to December 2013. . . . 17

2.3 Time series plots of the regional dispatch price and scatter plot of dispatch price again load for Queensland from 30 November 2012 to 7 November 2013. 21

3.1 Averaged half-hourly load over a day and averaged half-hourly load over a week in Panels (a) and (b) respectively, for Queensland over the period from 12th July 1999 to 27th November 2013. 32

3.2 Queensland load and temperatures, from July 1999 to December 2013. Solid line denotes a nonparametric regression fit with normal kernel and bandwidth 1. Dashed line is the ordinary least squares fit with the four temperature variables $\mathbb{C}_{1hd}$, $\mathbb{C}_{2hd}$, $\mathbb{H}_{1hd}$ and $\mathbb{H}_{2hd}$. The ranges of defined temperature variables in which they deviate positively from zero are indicated by the arrows. 35

3.3 Half-hourly MAPEs and overall MAPE for the prototype model, Equation (3.1). The overall MAPE is denoted as the solid horizontal line with its value indicated below. 36

3.4 The mean half-hourly forecast errors in days of a week over all weeks for the prototype model, Equation (3.1). 38

3.5 In Panel (a), the half-hourly MAPEs and overall MAPE for the prototype model (solid lines) in (3.1) are compared to the model with seasonal patterns (dashed lines) in the parameters given in (3.2). The overall MAPEs are shown as horizontal lines with the value for Equation (3.2) indicated below. In Panel (b), the mean half-hourly forecast errors in days of a week over all weeks from the prototype model (Equation (3.1), solid line) and the model with seasonal patterns in the parameters (Equation (3.2), dashed line). . . . 39

3.6 Estimated parameters and 95% confidence intervals (shaded areas) for weekly dummy variables in the model with seasonal patterns in the parameter (Equation (3.2)). On the left vertical axis, the deflections of the parameter estimates of Monday, Saturday, Sunday and other weekdays from Wednesday are denoted by dotted line with dots, dotted line with circles, dotted line with squares and solid lines respectively. On the right axis, the level of parameter estimates for Wednesday are plotted in dashed line. 41

3.7 The half-hourly MAPEs and overall MAPE for the mode with seasonal pattern in the parameters (Equation (3.2), dashed lines) and the model with the most recent load information in (3.3) (dotted lines). The overall MAPEs are shown as horizontal lines with the value of which for Equation (3.3) indicated below. 43

3.8 Forecast comparison on the half-hourly MAPEs for all the four models (solid line for Equation (3.1), dashed line for Equation (3.2), dotted line for Equation (3.3) and dash-dot line for Equation (3.4)) studied and the overall MAPE for the model using recursive system in (3.4) (dash-dot horizontal line with its value indicated below). 44

3.9 The half-hourly MAPEs of the one day ahead forecast produced by the proposed method (Equation (3.4) without temperature and special days, denoted by solid lines), the single equation double seasonal ARIMA (Equation (3.5), denoted by dashed lines), and the unconstrained intra-day cycles double seasonal HWES (Equation (3.6), denoted by dotted lines) from July 2002 to December 2013. The overall MAPEs are shown as the horizontal lines with the values indicated above. 47

3.10 The half-hourly MAPEs of 12-hour ahead forecast by Equation (3.4) (solid lines) and CEG (dashed lines) from July 2002 to December 2013. The overall MAPEs are shown as the horizontal lines. 49

3.11 12-hours ahead forecasts comparison of monthly MAPE between Equation (3.4) (solid line), Equation (3.4) without the future temperature (dotted line), CEG (dashed line) and the AEMO forecast (dot-dash line), from July 2012 to November 2013. 51

4.1 Half-hourly coverage ratios of the three forecasts, Model 1 (solid line), Model 2 (dotted line) and AEMO forecasts (dashed line) at quantile 0.1, 0.5 and 0.9. The evaluation period starts from 00:30 2014-11-24 to 00:00 2015-08-31. . 67

4.2 Half-hourly pin-ball losses of the three forecasts, Model 1 (solid line), Model 2 (dotted line) and AEMO forecasts (dashed line) at quantile 0.1, 0.5 and 0.9 in Panel (a), (b) and (c), respectively. The evaluation period starts from 00:30 2014-11-24 to 00:00 2015-08-31. 68

4.3 Half-hourly pin-ball losses of the three models compared, Model 1 (solid line), Model 2 (dotted line) and AEMO forecasts (dashed line) at quantile 0.1, 0.5 and 0.9. The left (right) column are pin-ball losses for positive (negative) residuals. The evaluation period starts from 00:30 2014-11-24 to 00:00 2015-08-31. 68

4.4 Histogram plots of residuals from randomly selected MCMC sample based on the first three-year moving window dataset at the 29th half-hourly interval of a day with quantiles at 0.1 (bottom row), 0.5 (middle row) and 0.9 (top row), respectively. The superimposed solid lines are fitted ALD density and non-parametric density obtained from Models 1 and 2, in left and right columns, respectively. 70

5.1 Panel (a) shows a time series plot of dispatch prices while Panel (b) shows a scatter plot of dispatch prices and loads. The data are for five-minute intervals in the Queensland region during the period 30 November 2012 to 7 November 2013, $T = 98784$ observations. 78

5.2 Scatter plot of the regional price differences between Queensland and New South Wales (Queensland price minus New South Wales price) and available import capacities on the inter-regional transmission lines for Queensland. The data are for five-minute intervals during the period 30 November 2012 to 7 November 2013, $T = 98784$ observations. 80

5.3 The price differences between Queensland and New South Wales plotted against inter-regional flows from New South Wales into Queensland. All the price differences, price differences without regional constraints, price differences without nodal constraints and price differences without any constraints are shown in Panels (a), (b), (c) and (d), respectively. 82

5.4 Time series plots of changes in dispatched quantities and changes in bid quantities for Queensland from 30 November 2012 to 7 November 2013, $T = 98784$ observations. 84

5.5 Time series plots of the regional dispatch price and constraints dummy variable for Queensland from 30 November 2012 to 7 November 2013, $T = 98784$ observations. 85

5.6 Time series plots of residuals obtained from ordinary least squares and bisquare M regressions, for the first 10,000 dispatch intervals. The level of residuals being displayed is restricted to be between -50 and 50 for a clearer presentation. 89

5.7 Estimated parameter values from quantile regressions of the form given in Equation (5.3) for 50 equally spaced quantiles from 0.01 to 0.99. 93

5.8 The price levels at 50 equally spaced quantiles based on the simulated price from the bootstrap procedure. The medians of the simulated quantiles of price with and without constraints are shown as dots and crosses, respectively, with the corresponding 95% intervals indicated as shaded area. 96

6.1 The changes of market cap price in the NEM. Dec 1998 is the starting date of the NEM. In July 2012, the cap price starts to be adjusted annually according to the inflation. 100

6.2 Panel (a), a times series plot of dispatch prices. Panel (b), a scatter plot of dispatch prices and dispatched quantities for Queensland, from 04:05, 30 November 2008 to 04:00, 24 June 2014. 105

6.3 The 288 five-minute bid curves for Queensland, 28 August 2013. 106

6.4 Scatter plots of system load versus the logarithm of price for Queensland in the months of June 2007 (top panel) and January and February 2009 (bottom panel). The dotted line represents the threshold value (natural logarithm of A$80/MWh) above which an irregular price event occurs. 107

6.5 Detailed bid curves in two consecutive dispatch intervals. Panel (a), (c) and (e) are the bid curve at 06:35, 28 Aug 2013 (before a price spike). Panel (b), (d) and (f) are the bid curve at 06:40 28 Aug 2013 at the time of a price spike. The dispatch prices and dispatched quantities are denoted by dotted horizontal and vertical lines with the values indicated beside. The bid curves are shown for three separate groups. The square boxes shown in Panel (a) and (b) are bids from generators with wide bid price range (with generator ID, generation capacity and fuel type: TNPS1 (443 MW, coal), TARONG#1 (365 MW, coal), TARONG#3 (365 MW, coal), STAN-1 (365 MW, coal), STAN-2 (365 MW, coal), STAN-3 (365 MW, coal), STAN-4 (365 MW, coal), GSTONE4 (280 MW, coal) and SWAN_E (370 MW, gas)). The circles shown in Panel (c) and (d) are bids from generators that only bid at the level above A$66.03/MWh. The remaining bids are shown as crosses in Panel (e) and (f). All bids are arranged in merit order. 111

6.6 Quantile regressions for the spans of bid prices of generators with ID STAN-3 (solid line), KPP_1 (dotted line) and MACKAYGT (dashed line). 95% credible intervals are shown as shaded areas. From Panels (a) to (f), the estimates for the parameter of constant and dummy variables defined after Equation (6.1) based on a grid of 50 equally spaced quantile from 0.01 to 0.99. 114

6.7 Illustrating the rebidding behaviour of base-load generators in Queensland between 06:30 and 07:00 on 28 August 2013. 116

6.8 Panel (a), time series plot of changes in bid quantity (ΔB_t). Panel (b), histogram of ΔB_t for all t. Panel (c), histogram for ΔB_t in the dispatched interval with a price spike above A$5,600/MWh. Panel (d) and (e), histogram of ΔB_t for the first and second dispatch intervals of a trading interval after a price spike (over A$5,600/MWh) was occurred in the one of the first four dispatch intervals of that trading interval. Panel (f), the cumulated changes in bids in the remaining dispatch starting from the second dispatch intervals after the spike to the last in the same trading interval. Panel (g), histogram of ΔB_t in the first dispatch interval after the trading intervals with the spikes. . 118

List of tables

3.1 The forecast comparison between the prototype model in (3.1) and the model with seasonal patterns in the parameters (Equation (3.2)). 40

3.2 The forecasting accuracy for the models studied, from Equations (3.1) to (3.4). 45

3.3 Summary comparison of 12-hours ahead forecast by Equation (3.4), CEG and the AEMO forecasts. 50

4.1 A summary of the forecast accuracy of the two proposed models and AEMO forecasts based on coverage ratio and pin-ball loss criteria. The evaluation period starts from 00:30 2014-11-24 to 00:00 2015-08-31. 66

5.1 Summary statistics for dispatch prices with or without transmission constraints, in Queensland from 30 November 2012 to 7 November 2013. . . . 85

5.2 Ordinary least squares, M and median regressions of the model in Equation (5.1). The dependent variable is the five-minute dispatch price in Queensland from 30 November 2012 to 7 November 2013, $T = 98784$ observations. The values in parentheses are t statistics. 88

5.3 Quantile regressions and M regression for the effect of transmission constraints with quantiles, $\tau \in \{0.5, 0.75, 0.9, 0.95, 0.99\}$. The dependent variable is the five-minute dispatch price in Queensland from 30 November 2012 to 7 November 2013. The values in parentheses are t statistics based on a robust estimate of the covariance matrix. 91

5.4 Kolmogorov-Smirnov tests of the null hypotheses of no constraint effects, constant constraint effects and positive constraint effect on price. Critical values are based on 10,000 block bootstrap replications with block size of 288. 95

6.1 Summary information for the regions in the NEM, from 1 June 2014 to 31 May 2015. The price information is based on regional dispatch prices. . . . 103

6.2 Dispatch prices, load, dispatched quantities and inter-regional out-flows of Queensland in a typical half-hour interval with a price spike. 108

6.3 Summary of the characteristics of three generators considered in the comparison. The sample period is from 1 Jan 2008 to 24 Jun 2014 (681,696 of total dispatch intervals). The number of dispatched intervals for each generator is calculated as the number of intervals where the dispatched price exceeded the minimum bid price (with positive available capacity) of the generator. . 113

6.4 Quantiles ($\tau = 0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95$) estimates of the distributions of ΔB_t in the dispatched intervals of all, around price spikes and of trading intervals with potential rebidding. (95% credible intervals based on the posterior draws are reported in brackets) 120

Chapter 1

Introduction

1.1 Motivation

Electricity, as a main source of energy, fulfils needs which are fundamental to the normal functioning of our society. The corresponding markets, electricity markets, are undoubtedly one of the most important type of commodity markets that provides support to economic growth. To secure the supply of electricity and ensure the normal operation of the market, most electricity markets around the world were heavily regulated before the 1990s. In these markets, the generation, transmission, distribution and retailing assets were all owned entirely by governments, and the price of electricity was held fixed. In this way, the development of the corresponding infrastructure was centrally planned in ordered to match the growing demand, and the financial risk involved in the trading of electricity was eliminated for the end users.

However, a well-functioning market that can best support a nation and its economic growth is one which not only can provide security of supply, but also operates efficiently in a sense that the price of the product reflects its underlying cost. In this way, all market participants can benefit from having access to a freely competitive and low cost market. It is

for this reason, a wave of deregulation in the electricity markets world wide started in the early 1990s.

In the case of Australia, the electricity sector was vertically monopolised by the government across generation, transmission, distribution and retailing. Following the deregulation of the electricity market in the United Kingdom (1990), the National Electricity Market (NEM) of Australia commenced in December 1998 after seven years of preparation. During the period of preparation from 1991 to 1998, the design of NEM as a deregulated market was developed, inter-state transmission networks were built between some of the states, and restructuring of the government owned assets took place. The restructuring involves disaggregation, corporatisation and privatisation of the assets both vertically and horizontally.[1]

As of 2015, full retail competition was achieved in all jurisdictions of the NEM except Tasmania. Note, however, that the introduction of competition in generation is slower than that in retailing, with 67% of the generation capacity in Queensland, and all capacity in Tasmania still state owned. Like the generation market, the privatisation of transmission market also varies by regions, with the Queensland and Tasmania networks fully owned by the government, New South Wales network partly owned by the government, and all transmission network in Victoria and South Australia, and part of inter-regional networks privately owned.[2]

1.2 The structure of the NEM

Like other competitive commodity markets, the general structure of the NEM can be represented in a simplified market equilibrium framework, in which the price, demand and supply of electricity are jointly determined. On one hand, the intersection of the demand and supply curves determines the equilibrium price. On the other hand, the demand and supply of electricity are affected by the equilibrium price through the behaviour of consumers and generators in response to the price. However, beyond this simplified framework, many complications can arise due to the unique characteristics of electricity markets. Chapters

[1] See, Australian Energy Market Commission (2013) for a more detailed NEM history.
[2] See, Australian Energy Regulator (2015).

3 and 4 focus on the problem of demand forecasting, which is of central importance to the dispatch planning of the market operator due to the non-storable nature of electricity and the real-time balanced equilibrium. In Chapter 6, the issue of strategic bidding and rebidding of generators arise from the supply side due to the constantly evolving deregulation process and the market rules. From a more general point of view, Chapter 5 analyses the behaviour of the equilibrium price with special attention paid to the effect transmission constraints. Consequently, the detailed examinations of these issues necessarily requires in-depth knowledge about the market structure and the physical characteristics of electricity markets and the NEM in particular.

The NEM is one of the world's largest deregulated electricity markets which comprises the regions of New South Wales (NSW), Victoria (VIC), Queensland (QLD), South Australia (SA) and Tasmania (TAS). These five regional markets are further subdivided into smaller areas called nodes. A transmission network links the generators with load centres covering around 4,500 km while the distribution network, which transports electricity from points along the transmission network to final users, is 17 times larger at around 750,000 km. A feature of the NEM is that in eastern and southern Australia the different regions are fully interconnected. QLD and NSW are connected by two interconnectors, (QNI and Directlink) with total capacity of around 400 Mega Watts (MW) into QLD and 1,100 MW into NSW; VIC and NSW are connected by one interconnector with capacity of around 1,900 MW into VIC and 3,200 MW into NSW; Murraylink with capacity around 200 MW in both directions connects VIC and SA; and Basslink (around 500 MW) connects VIC and TAS. Based on this infrastructure, the trading between around 300 registered generators with total installed capacity of around 48,000 MW and nine million customers is operated as a pooled market under the supervision of the Australian Energy Market Operator (AEMO).

Since the Queensland region is given special attention in this thesis, it is useful to spell out the Queensland network in a little detail. Queensland is connected to the rest of the NEM by two transmission lines, QNI (Queensland New South Wales Interconnector) and Directlink, that cross the border between Queensland and New South Wales. The import capacity of QNI and Directlink into Queensland is around 150 MW and 20 MW respectively.

The exact capacity varies with many factors that are related to the physical characteristics of the network and the market in general. Within Queensland itself, the main transmission lines are located along a long strip of the coastal area where the majority of the population is located. This configuration means that any occurrence of transmission line outage in the region is likely to separate between the north and the south of the region and completely isolate the north from the rest of the market.

Wholesale trading in the NEM is conducted as a pooled spot market where supply and demand in each node are instantaneously matched through a centrally-coordinated dispatch process managed by the AEMO. In this market, generators are allowed to bid generation capacity within 10 price bands which fall between the floor price of -1,000 AUD/MWh and the cap price of 13,500 AUD/MWh. The initial bids of generators are made daily for the next day, however, a unique feature of the NEM is that despite their initial day ahead bids, generators can change or shift their offer capacity at any time five minutes before the actual dispatch. These bidding rules enable generators to adjust their trading strategy and maintenance plan flexibly in response to the current market conditions.

The central dispatch algorithm uses the current bid information to schedule generation in order to satisfy the demand for electricity at all nodes at minimum cost with the cost measured by the bid prices of generators. In scheduling the dispatch strategy, the transmission networks and its physical limits play crucial roles. That is because the dispatch strategy not only utilises the generation capacity located at the same location to fulfil the demand, but also the capacity at other locations through the transmission network. If the cost were different on the two ends of a transmission line, the higher cost at one end is the consequence of dispatching generator in the area with higher bid price. Given there is extra transmission capacity available on the line, the overall cost measured by bid price can be reduced by dispatching generation resource at the end with a lower bid price and transferring the electricity into the end with higher price. Consequently, with the aid of transmission network, the dispatch strategy is formed based on minimising of overall cost of fulfilling the demand in the whole market subject to certain physical constraints of the network.

Among the physical constraints in the network that a dispatch strategy has to satisfy, the most prominent one is transmission constraints. Transmission constraints refers to the physical limit of the transferring capacity of transmission lines, which are subject to various technical reasons such as temperature and frequency controlling. Without transmission constraints, dispatch algorithm can utilise generation capacities at any location in the network, consequently leading to the dispatched bids of generators to have bid prices very close to each other in the whole market. On the other hand, the occurrence of transmission constraints in the network inhibits the flow of electricity across the market, and the result can be differing market conditions and potentially large differences between generation costs among the locations.

In addition to transmission constraints, the dispatch algorithm has also to consider the synchronisation rate and ramp-up rate of generators. The ramp-up rate is the speed of increase in generation capacity for an in-use generator that has not yet reached its maximum capacity, while the synchronisation rate is the time a generator that is not currently dispatched requires from receiving a dispatch instruction to start injecting electricity to the grid. The ramp-up rate is generally a lot faster than the synchronisation rate for any kind of plant. Gas-fired marginal generators, acquired over time in order to mitigate the effects of sustained periods of abnormally high prices, usually take about 5 to 30 minutes to come online. Whereas a generator that is already generating electricity but not at capacity, can increase its output by potentially more than 200 MW per minute. What this means to dispatch algorithm is that even extra generation capacity is available based on the current bidding information. In order to balance the real time load changes in the network, the dispatch algorithm has to choose the one not only with the lowest bid price but also capable of responding in a timely manner. In cases, to prevent blackout, generators located in other locations and with high bid price have to be used to replace the local slow response generators with lower bid price.

By taking into account those major physical constraints and other minor ones, the equilibrium spot price of electricity is determined as the outcome of the dispatch process. The preliminary nodal price is calculated for each node in five-minute intervals, based on the real-time nodal equilibrium conditions. The five-minute regional dispatch price, subject to

minor adjustments, is given by the highest nodal price recorded in the region. For settlement purposes, a half-hourly regional trading price is quoted, which is the average of the six regional dispatch prices in the half-hour trading interval.

Overall, the physical structure of the market including generators and load centres located in different locations, the transmission and distribution networks, along with the bidding, dispatch and price setting rules forms the general framework of the deregulated market, which all market participants must follow.

1.3 The research questions

As can be seen from the previous sections, the adventures into deregulation with fundamental restructuring and creation of both new and complex system, can bring significant challenges to both the policy makers and all other market participants. For example, the introduction of competition and privatisation means that private investors in generation assets, transmission and distribution networks, and retailing can operate as individuals in the market with the goal of profit maximisation. To achieve this goal, the strategic decision involved in their daily operation naturally requires adequate knowledge of future market conditions. As the main indicator of the demand side condition, the forecast of demand becomes important to not only the regulators in terms of central planning just as in regulated market, but to the decision making of all other private market participants. This gives rise to a new significance and need for accuracy in demand forecasts.

Not only old issues become significant due to the deregulation, the structural reform involved in the deregulation process also creates new challenges. For example, the price of electricity in a free market must be flexible in order to reflect the fundamental cost and the changing conditions of the market. Consequently, a fixed price cannot be used, instead, certain flexible price setting rules are required. This raises the question of how prices move according to the various changes in market conditions, especially in respect to those unique characteristics of electricity markets, such as transmission constraints.

Moreover, as the result of the free price movement, market participants react accordingly, and their behaviour becomes important to the market regulator in terms of policy making. For example, the bidding behaviour of generators can play an important role in determining whether the current market is competitive or not. If the bid behaviour of an individual generator can lead to sustained price changes, it could reflect the market power of that generator. Whereas, if under the current rule, a price event can lead to market participants behaving strategically at the cost of others, and it may indicate the need for improvements in either the current bidding policy or price setting rules.

Overall, the success of the deregulation in an electricity market requires not only an adequate regulatory environment to be set up for maintaining the balance between the security of electricity supply and the freedom allowed in the market for competitiveness, but also the smooth operations of all participants in the market. In order to achieve this, a thorough understanding of the nature of deregulated electricity markets and the major issues involved is required. Consequently, it gives rise to the following four specific research questions that form the basis of this thesis:

Day ahead point forecast of electricity load. Point forecast of electricity load has long been the centre of the forecasting literature. As it progresses, models for load forecasting have become increasingly complex, computational intensive and flexible with attention paid to accommodating the non-linear response of load to many well-known driving factors. Based on this phenomena, a key research question is whether higher accuracy is achieved by increasing flexibility and complexity of models, or if there exists a simpler model which, with careful implementation, could compete or even outperform the more complex ones.

Quantile forecasting of electricity load. Quantile forecasts of load can represent the potential variation of future load, and provide valuable information for dispatch planning and strategic decision making. Consequently, the second research question is how to formulate the load forecasting problem in the context of quantile regression, and how to produce quantile forecasts of load with satisfactory forecasting accuracy.

The behaviour of electricity price and effect of transmission constraints. The highly variable electricity price is a key feature in a deregulated electricity market, whose behaviour can be regarded as a main indicator of the competitiveness of the market. This leads to the third research question, namely how the price in a deregulated electricity market behaves in response to changes in market conditions, and in particular, what the effect is of transmission constraint on prices.

Strategic bidding and rebidding behaviours of generators. In responding to the free movement of price, the behaviour of market participants can be useful for determining the successfulness of the regulatory aspect of market deregulation. Strategic behaviour by generators that is derived from market power can be particularly harmful to the operation of the market and contradict the intention of market deregulation. Therefore, the final research question is whether there exists strategic bidding behaviour by generators in the context of NEM, and how it relates to the various extreme market conditions caused by the physical characteristics of electricity market.

1.4 The structure of the thesis

The four research questions of this thesis can be thought of as arising from the different aspects of a deregulated electricity market. The problems of load forecasting, price behaviour and strategic bidding reflect, respectively, the demand side, general equilibrium and supply side of the market. In preparing for the detailed study of these research questions, Chapter 2 provides an overview of the corresponding literature.

Chapter 3 proposes a model for one day ahead point forecast of load. The formulation of the model starts with a basic autoregressive moving average structure. Then the basic model specification is expanded following a study on the effect of the main driving factors of load, such as seasonality, temperature and special days. In particular, the seasonality of load is given special attention in the final version of the proposed model, which forms the key distinction of the proposed model from its predecessors in the forecasting literature. It turns out that allowing for a distinct weekly pattern in the coefficients governing one-

day lagged load is a crucial advance on previous work. The efficacy of this innovation in dealing with seasonality is demonstrated by comparing with two traditional ways of dealing with seasonality, namely the double seasonal ARIMA and the double seasonal Holt-Winters exponential smoothing approach (Gould et al., 2008). Incorporating the proposed refinements into a multiple equation model, the superior forecasting performance of the final chosen model is shown by comparing with the multiple equation model of Cancelo et al. (2008) and a semi-parametric approach used by the AEMO.

In Chapter 4, the attention focuses primarily on how to produce quantile forecast of load using quantile regressions, and how to maintain the computational tractability in the context of quantile load forecasting while being able to accommodate a large number of covariates, which were found to be useful for load forecasting in Chapter 3. As a result, two quantile load forecasting models are formulated in a Bayesian framework. They differ mainly in the type of underlying residual distributions. The forecasting results of the proposed models are compared with each other and also with an industry standard reported by AEMO. Overall, the proposed models perform competitively in comparison with the AEMO forecast, and offer interval forecasts of load that can reflect the uncertainty of the actual load around the forecast and the risk involved in the decision making process of the market participants.

Moving to the study on the price behaviour and the effect of transmission constraints in Chapter 5, the experience gained from quantile load forecasting (Chapter 4) allows the investigation in this part to be based not only on identifying the mean effects of transmission constraints on prices but also the quantile effects. The main approach adopted in this chapter is to use quantile regressions to identify the effect of transmission constraints on slices of the price distribution. The result indicates that transmission constraints are an important cause of short-term price variation, a result that emphasises the importance of which to any model that seeks to forecast regional electricity prices in an integrated electricity market.

Continuing on the study of price, in Chapter 6, the behaviour of the market participants in response to price movements, more specifically, the strategic bidding and rebidding behaviour of generators is studied in connection with extreme price events. This is achieved using detailed dataset, in which interesting irregularities in the movement of price, inter-regional

flow, load and dispatched capacity, can be related directly to the interaction of the physical characteristics of electricity markets and the strategic behaviour of market participants. In particular, it is shown there exist strategic bidding and rebidding behaviour of generators in the NEM, and this behaviour is connected directly with the irregularities in the movement of price and the physical characteristics of the market. In addition, it is also found that the irregularities in the movement of price are exacerbated by the current policy rules governing bidding behaviour in the NEM. That is, the occurrence of counter-price flows can be related directly to the rebidding behaviour of generators after extreme price event. As a result, policy options for dealing with these problems are proposed at the end.

1.5 Key contributions

Overall, this thesis makes significant contributions to our understanding of the operation of deregulated electricity markets. Detailed investigations into three major aspects of electricity markets, namely, demand, supply and general equilibrium are undertaken. The findings derived from these investigations advance our current knowledge of deregulated electricity markets as a whole.

On the demand side, forecasts of future electricity load play central role in the dispatch planning of the market operator. Two models for load forecasting are proposed in this thesis. Despite having a simple modelling framework, the first model pays special attention to the interaction between seasonal pattern and achieves a superior forecasting performance in comparing to various state-of-the-art forecasting alternative and the industrial standard. The second forecasting model proposed complement the first by producing quantile forecast of load, which entails the uncertainty about the future load and decision risk involved. On the supply side, the bidding and rebidding behaviours of generators is of central concern to the market regulator. The successfulness of the market deregulation and the effectiveness of the current market rule in promoting competition. The bid and rebidding behaviour revealed provides vital information about the successfulness of the market deregulation and the effectiveness of the current market rule in promoting competition. From the aspect of the

general equilibrium of an electricity market, The price behaviour as the result of the general equilibrium and its relation to the fundamental characteristics of the market found in this thesis also allows market participants to better manage the price risk involved in their daily trading activities.

In Chapter 3, the multiple equation load forecasting model proposed pays particular attention to the interaction between daily and weekly load patterns. Probably the most important distinguishing factor in the proposed model relative to others in the literature is the flexibility built into the influence of load from the same half hour on the previous day. Allowing the strong weekly pattern to interact with the daily pattern in coefficients on lagged load yields important improvements in short-term forecast performance. Another innovative dimension of the model is the use of the inherent recursive structure of the model to capture the intra-day load correlation. The effectiveness of the proposed approach on modelling the seasonal features of electricity load is demonstrated by comparing with two popular alternatives, double seasonal ARIMA and Holt-Winters exponential smoothing. Despite these modifications to the preferred model, it remains linear in parameters and can be estimated equation-by-equation by ordinary least squares. Overall, the forecasting performance of the preferred model is impressive and significantly out-performs two benchmarks with which it is compared. In particular, the model improves on the mean average percentage error of 12-hour ahead forecast reported by the Australian energy market operator by about a third. For the entire 11 year period, the model returns a mean average percentage error of 1.36% on half-hourly day-ahead forecasts, a figure that is lower than most (if not all) comparable average error statistics reported in the literature.

Replacing the least squares procedure for point forecasting of load in Chapter 3, Chapter 4 presents two quantile forecasting models in a Bayesian framework with the residual distribution being set to follow either a asymmetry Laplace distribution or a non-parametric residual distribution. The key feature of this approach is the use of a Gibbs sampler that permits the sampling of parameters to be broken into two batches, one for observed covariates, one for recursively defined covariates. Moreover, due to the computational advantage of the algorithm for sampling the parameters of observed covariates, a large number of

parameters can be accommodated, which is crucial in the context of quantile load forecasting for accounting many important covariates such as temperature and special days, and for achieving satisfactory forecasting performance. Overall, in addition to the point forecast model of load of Chapter 3, the proposed quantile load forecasting models can provide the participants in an electricity market with additional information about the variation of future load, and reflect the risk involve in their strategic decision making process and operation.

In the study on electricity price, despite significant advances in the understanding of deregulated electricity markets and in econometric methodology insofar as it pertains to modelling electricity prices, the covariates traditionally used to explain variations in electricity prices have remained relatively static. These covariates, such as load, temperature and reserve margins, are smoothly-varying leaving only the lagged price capable of capturing the rigidity and volatility of spot prices. By contrast, Chapter 5 contributes to the literature by investigating the role of transmission constraints as a fundamental contributor to the variability of electricity prices. In order to identify the effects of transmission constraints, high frequency five-minute data recently made available by AEMO is employed. Moreover, the results reported in this chapter are based mainly on quantile regressions, which are used to ensure robust estimation and inference in the presence of extreme price outliers. It is found that transmission constraints contribute significantly both to the level and variability of price. Consequently, the performance of a price forecasting model is likely to be improved by incorporating information on transmission constraints. It is also demonstrated that the presence of constraints is a necessary condition for the occurrence of extreme prices. One result of particular interest is that price is explosive in the upper tail of the price distribution, a result driven entirely by the presence of constraints.

In Chapter 5, although transmission constraint is shown to be a necessary condition for the occurrence of extreme prices, there is clearly a need for identifying the other fundamental drivers of extreme prices given that transmission constraints are not sufficient conditions for these events. Chapter 6 continues on this exploration using a large and detailed dataset that encompass much useful market information, such as price, dispatched load, demand, inter-regional flows, bid information of every generators, etc. at the same five minute frequency.

Based on this detailed dataset, the physical constraint of ramp-up rates and synchronisation rates of generators are found to play a crucial role in the occurrence extreme price events. This leads to the key contribution of Chapter 6, the existence of bidding and rebidding behaviour of generators in the NEM. More specifically, it is found that the strategic bidding of base generators in the form of bid splitting takes the advantage of the temporary power enabled by the ramp-up rates of which in relative constraint regions, and contribute directly to the extreme level of price events. And then the rebidding after extreme price events enabled by the current bidding rules raises a series of substantial but short changes in market conditions that are harmful to the operation of the market. In particular, it is found that the rebidding of generators after extreme price events contributes directly to the occurrence of counter price flow.

Chapter 2

The relevant literature

2.1 Load forecasting

Accurate demand forecasting is important to the participants in any of the commodity markets. It is even more so in the context of the electricity market. This importance arises from the physical nature of electricity markets. To explain, electricity is the only commodity that can be generally considered as non-storable economically in large scale. As such, the market equilibrium has to be balanced in real time with the generation of electricity to be flexible in order to match the constantly changing demand. Consequently, accurate load forecasts carry useful demand side information, and enable the market operators and other participants to prepare in advance for changes in market conditions.

The focus of the first part of this thesis is short-term load forecasting up to one day ahead. In recent years, due to the deregulations of the electricity market world-wide, short-term forecasting has became increasingly important. The reasons for this increasing importance of accurate short-term load forecasting differ for each of the players in a deregulated electricity market. From the perspective of the market operator, short-term forecasting is crucial to the scheduling and dispatch of generation capacity; for the electricity generators, the strategic choices involved in bidding and rebidding of capacity depend critically on load forecasts; and for the electricity retailers, load forecasting affects decisions about the balance between

hedging and spot acquisition of electricity. For these reasons, short-term load forecasting remains a problem of central interest and one which has generated a large literature.

The fundamental drivers of electricity load that has been identified in the literature includes seasonality, temperature and special days (Srinivasan et al., 1995; Pardo et al., 2002; Kim, 2013). In the case of the seasonality, for example, Figure 2.1 plots the average half-hourly load over a day and average half-hourly load over the period of a week using the Queensland data with the average taken over the entire sample period from 12th July 1999 to 27th November 2013. Diurnal and weekly patterns, both well documented features of electricity load (Engle et al., 1989; Harvey and Koopman, 1993; Taylor, 2010), are clearly evident. Load picks up very quickly between the hours of 06:00 and 08:00 from the overnight low and remains high during the daylight hours. The daily peak in the load profile usually occurs at 18:00 before tailing off once more. The weekly pattern in load is also quite pronounced with a regular load profile evident from Monday through Thursday, but with significant differences on Friday, Saturday and Sunday.

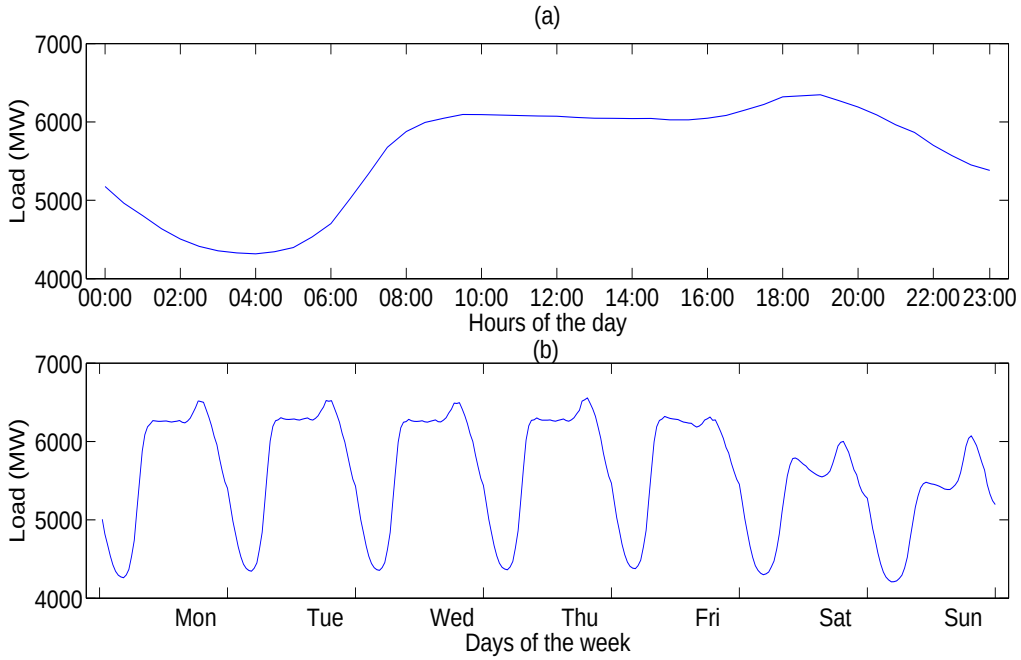

Fig. 2.1 Averaged half-hourly load over a day and averaged half-hourly load over a week in Panels (a) and (b) respectively, for Queensland over the period from 12th July 1999 to 27th November 2013.

For temperature, Figure 2.2 demonstrates the non-linear nature of load in response to temperature using the Queensland data. It can be seen that the load peaks at both high and low temperature ranges when the need of cooling and warming rise, and reaches the lowest level in the middle temperature range. Also, as the temperature extends into the more extreme regions, the load level tends to remain flat due to the limit on the maximum installed capacity of temperature controlling devices.

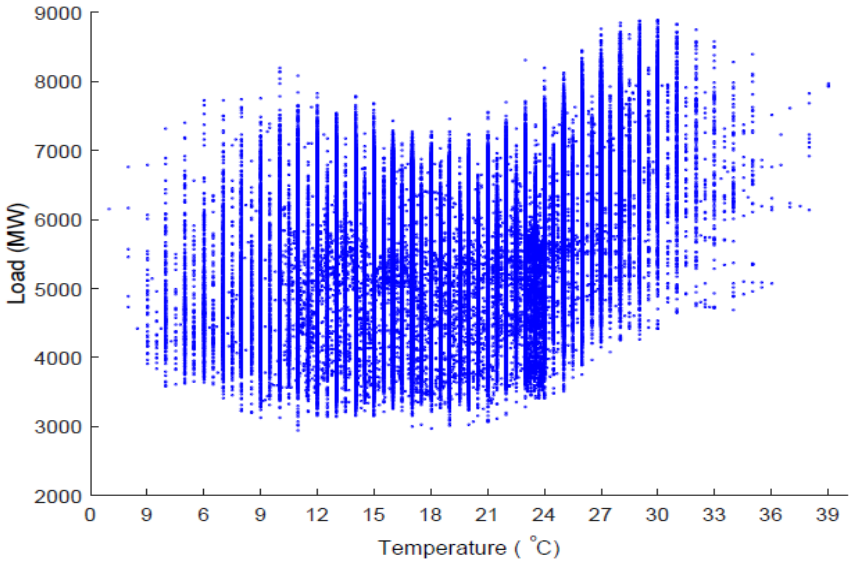

Fig. 2.2 Queensland load and temperatures, from July 1999 to December 2013.

Based on these well identified drivers of load, statistical models for short-term load forecasting fall very naturally into three main categories. The first of these is single equation time series models, which model the trajectory of load using traditional time series methods (Hagan and Behr, 1987; Darbellay and Slama, 2000; Taylor and McSharry, 2007; Gould et al., 2008). The efficacy of this approach derives from the strong seasonal patterns in electricity load demonstrated in Figure 2.1. In dealing with the seasonality, a popular choice is to use the Autoregressive moving average (ARMA) type specification of Box et al. (2015). In this approach, the forecast of load is produced based on a linear combination of lagged load and residuals from previous days and previous weeks (for example, Kim, 2013), which are aligned with the observed seasonality in Figure 2.1. However, an important piece of

information missing in the basic ARMA specification is the intra-day pattern. That is, when estimating the parameters of the model, the loads on different hours of a day are treated without distinction, and the parameter obtained is the average impact over all the hours of a day. This observation leads to an expanded version of the ARMA specification, termed as periodic ARMA, in which, every parameter is redefined use polynomial functions (Taylor and McSharry, 2007; Amaral et al., 2008).

The second, and probably the current method of choice for practitioners, is the neural network approach in which the trajectory of load is modelled semi-parametrically using basis functions with emphasis on the non-linearity of load (Park et al., 1991; Zhang et al., 1998; Hippert et al., 2001; Taylor and Buizza, 2002). In this approach, the model specification is allowed to be data dependent, which adds arbitrary flexibility to the model specification not only for the part accounting for temperature effect as shown in Figure 2.2 but for the effects of all other covariates.

The third type of statistical model used to forecast load is the multiple equation time series class. These models have enjoyed some popularity in the literature but their influence has waned in recent years. In this approach, each period of the day (usually each half hour or hour) is treated as a separate forecasting problem with its own equation (Peirson and Henley, 1994; Ramanathan et al., 1997; Espinoza et al., 2005; Soares and Medeiros, 2008; Cancelo et al., 2008). In this way, it provides an alternative to the single equation periodic ARMA approach for dealing with the intra-day seasonality of load. It is this class of model that forms the basis of the proposed model in Chapter 3.

In regard to these various models proposed for short-term load forecasting, however, a common characteristic is that they all focus on producing point forecasts that are as close as possible to realised loads. Consequently, if only relying on these point forecasts in the decision making process of market participants, the risk involved could be hidden due to the uncertainty about the variation of future load around the point forecasts. To represent the uncertainty of future load, a relevant approach is to forecast the maximum demand in a day or a week (Engle et al., 1992; Hyndman and Fan, 2010), which is still a point forecast but carries information about the upper end variation of load within a certain period. Alternatively,

simulation procedures can be used. Here, the simulated covariates or residuals are used to construct interval predictions of load (McSharry et al., 2005; Fan and Hyndman, 2012). Furthermore, quantile regression has also been used on a set of point forecasts produced by a group of point forecast models (Liu et al., 2015), thus providing a sense of uncertainty arising from the different specifications of various point forecast models. To extend these ideas and also provide a more complete picture of the risk involved, it is natural to forecast the interval or quantiles of future load directly using quantile regressions with the quantiles being set at multiple points.

Although the application of quantile regression to forecasting the quantiles of various economic variables is not new (Taylor, 2007; Gerlach et al., 2011), it has rarely been used directly on forecasting load. Instead, quantile forecasts of load are normally produced based on simulating or bootstrapping residuals or historical observations (Fan and Hyndman, 2012). The reason for this phenomenon is that the classic quantile regression is computationally intensive even for models with a small number of parameters, whereas in load forecasting, complicated seasonal patterns, the effect of temperature and special days usually require a large number of parameters in order to achieve satisfactory forecasting accuracy. To overcome this issue while taking into account a large number of covariates, López Cabrera and Schulz (2014) used the approach of dimension reduction with principal components in the context of classic quantile regression framework. However, a well-known problem with using principal components is that small variations in the covariates do not necessarily imply low impact on the dependent variable. Consequently, important inferential information may be lost when using principal components, and further development in quantile forecasting of load may be required for better forecasting performance. Following this direction, Chapter 4 is focused on forecasting quantiles of load.

2.2 Price behaviour

The load forecasting problem on the demand side of the electricity market is important because it provides important information for the supply side to prepare for maintaining

the real-time equilibrium of the market in future in terms of dispatch planning and strategic decision making. However, from the general point of view of the market equilibrium, the single most important indicator is the price of electricity. In contrast to the fixed pricing rule in regulated electricity markets, a constantly changing price depending on the real-time market equilibrium is a key feature of the deregulated market. This variation in the price gives rise to an important question of how the price behaves in response to the changes in market conditions.

In general, the characteristics of the dynamic of electricity prices in deregulated markets that have been identified in the literature include, seasonality, jumps, and mean reversion. Similar to the seasonality of load described in the previous section, the seasonality of price follows the same daily, weekly and annual patterns (for example, Pindoriya et al., 2008; Bowden and Payne, 2008). Whereas extreme jumps and mean reverting behaviour of electricity price are unique, and have a great impact on the formulation of price models in the literature (Huisman and Mahieu, 2003; Cartea and Figueroa, 2005; Higgs and Worthington, 2008; Bhar et al., 2013). The extreme jumps include both negative and positive jumps, with a negative one normally followed shortly after by a positive one, thus resulting a short lasting price spike. In terms of mean reversion, it generally represents the process of price adjustment, which allows the price to gradually return to the level representing the fundamental cost.

To give a general view of the price dynamic faced in the NEM, the first panel of Figure 2.3 shows the Queensland regional dispatch price for the period from 30 November 2012 to 7 November 2013. A striking feature of the price movements that can be seen is the occurrence of short lasting extreme price events. In the period displayed, both the negative prices at near -1,000 A$/MWh and extreme positive prices at the cap price level are found. These correspond to the jump behaviour of the price. Beside the jumps, it can be seen the price tends to vary in a less volatile way at very low level, at which the price served as a reflection of the fundamental cost. In addition, the scatter plot of dispatch price and load for the same sample period in the second panel of Figure 2.3 shows that abnormal price events occur over a wide range of load and not merely at very high levels where the system could reasonably be characterised as being at capacity. To summarise, the irregularities of the price movement in

both level and in correspondence with load further raises the question of what are the causes of the irregular price events?

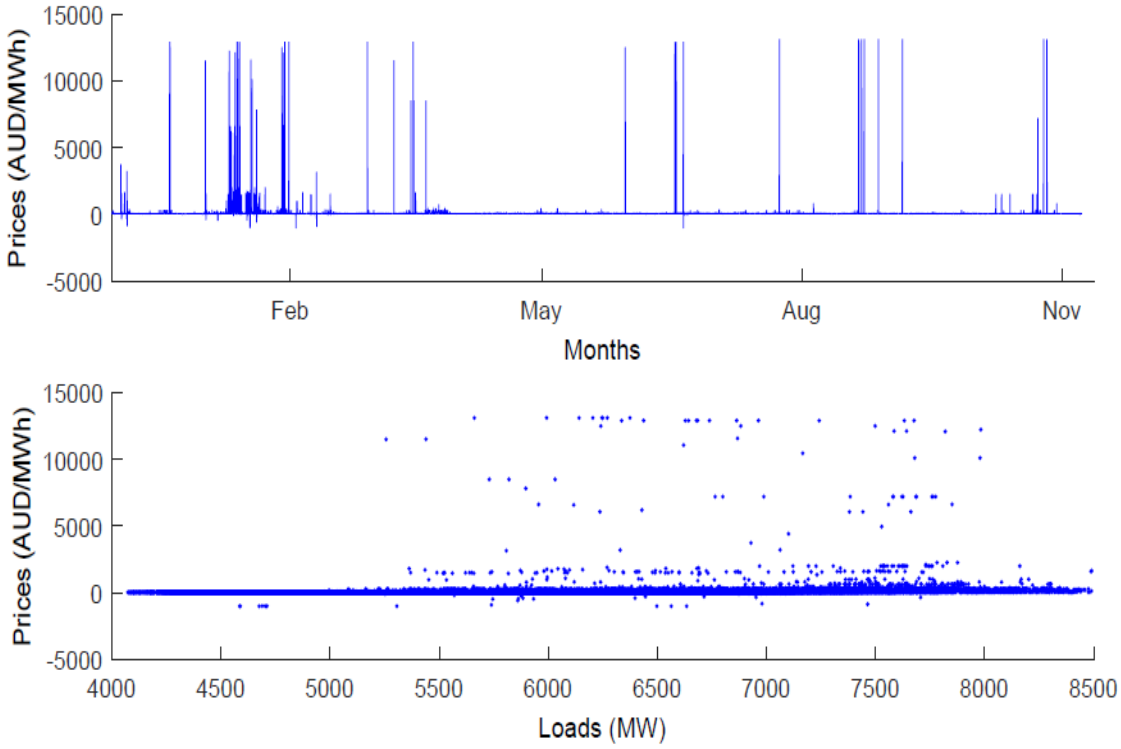

Fig. 2.3 Time series plots of the regional dispatch price and scatter plot of dispatch price again load for Queensland from 30 November 2012 to 7 November 2013.

From the perspective of a regulator, the movement of price reflects directly whether the market deregulation has resulted in the intended consequence of achieving higher market efficiency by having the price to be representative for the fundamental cost. For other participants in the market, the understanding of price behaviour contributes directly to accurate forecasting of it, and thus aids in managing the risk involved in trading activities. For these reasons, the behaviour of short-term electricity prices and the relevant forecasting methodology have attracted a significant amount of attention in the literature (see Weron, 2014, for a comprehensive survey).

In general, short-term price forecasting models can be categorised into three main categories. *First*, Stochastic models focus on modelling the trajectory of price by treating

it as a stochastic process. The two main types of stochastic models are: jump-diffusion models and regime switching models. The jump-diffusion models focus on capturing the main property of the price evolution, namely jump and mean-reversion (Cartea and Figueroa, 2005; Benth et al., 2007; Albanese et al., 2012; Bhar et al., 2013). In regime switching models, similar behaviours are considered by using different regimes for the states of the market with transitioning probabilities between each regimes (de Jong and Huisman, 2003; Karakatsani and Bunn, 2008; Eichler and Tuerk, 2013).

Second, time-series models include various autoregressive moving average type models (Contreras et al., 2003; Cuaresma et al., 2004; Cruz et al., 2011), threshold models (Robinson, 2000; Weron and Misiorek, 2008) and GARCH type models (Knittel and Roberts, 2005; Garcia et al., 2005; Huurman et al., 2012), that are mostly similar to those used in load forecasting literature. In the formulation of times series model, while seasonality of price is addressed similarly to that in load forecasting literature (Misiorek et al., 2006; Karakatsani and Bunn, 2008), other important drivers of price are also considered, in particular, weather (Huurman et al., 2012) and reserve margin (Zareipour et al., 2006; Anderson and Davison, 2008).

Third, Machine learning type models model the price by using various non-linear or non-parametric frameworks, which can provide arbitrary flexibility to fit the data. Neural networks (Yamin et al., 2004; Mandal et al., 2006; Areekul et al., 2010) and support vector machine (Sansom et al., 2002; Zhao et al., 2008; Chaâbane, 2014) are the two main types applied to forecasting electricity price.

Despite the various models and the driving factors considered in the literature, however, the price studied so far has either been restricted to one locational price for one area with explanatory variables that only represent the market conditions in that area, or an overall price level that is recorded prior to the change to locational pricing (Huisman and Mahieu, 2003; Cartea and Figueroa, 2005; Mount et al., 2006; Kanamura and Ohashi, 2008; Mount and Ju, 2014). This is in sharp contrast to the institutional setting, in which locational pricing (Ding and Fuller, 2005; Aderounmu and Wolff, 2014) that sets differential prices for smaller areas based on the local equilibrium conditions, are implemented in most deregulated markets.

Consequently, much of the existing literature on modeling prices ignores the important role of the transmission infrastructure in enabling equilibrium to be maintained at all locations on the grid. At the very least, a model of electricity prices must explicitly take into account transmission constraints which constitutes important information about the local state of the market for which forecasts are generated (Douglas and Popova, 2011; Burnett and Zhao, 2015).

To emphasise the important effect of transmission constraint on electricity price, Chapter 5 of this thesis explores the effect of transmission constraints on short-term electricity price variation in the Queensland region of the NEM, using detailed data about the operation of the NEM made available by the AEMO.[1] These data include five-minute dispatch reports containing the regional dispatch prices, dispatched capacities, loads, interregional flows and nodal constrains together with bid reports that provide information on the bidding process for the previous day. The use of five-minute data on dispatch prices, instead of the half-hourly trading prices commonly studied in the literature, has the advantage that they are the actual prices produced by the dispatch algorithm and therefore preserve the potential causes of electricity price variations without contamination by averaging. In this way a clearer identification of the cause of variations in trading prices is facilitated. Another innovative aspect of the dataset is the inclusion of comprehensive bid information, which contains the bids from all generators in the market at the same five-minute frequency. The bid information enables the construction of a variable, the changes in bid quantities, which represents the changes of the supply side market condition and is used in the regression analysis in order to facilitate the identification of the transmission constraints effect.

2.3 Strategic bidding and rebidding

In studying the effect of transmission constraints on electricity prices, a feature of the electricity market emphasised is the role of the transmission grid in an integrated electricity market. Because all locations within a region must be in equilibrium, any transmission constraints on

[1] See www.nemweb.com.au

the grid can lead to the isolation of some areas within the region and mismatches of local supply and demand. Consequently, in order to achieve a highly adaptive supply side which can respond effectively to the real-time market changes, generators in the NEM are allowed to adjust their bids at any time up to five minutes before the actual dispatch.[2]

The flexible bidding rule not only allows the supply side to adjust quickly and thus help to ensure the maintenance of real-time equilibrium, but also gives market participants a high degree of freedom in conducting their trading strategies, which theoretically should promote free competition in the market. However, flexible bidding may also lead generators to bid strategically with the undesirable consequences of enabling regional prices to deviate from the fundamental cost of generation, or to take advantage of temporary extreme market conditions such as abnormally high prices, caused by the physical nature of the market.

In order to examine the existence of strategic bidding behaviour of generators on supply side of the market, it is worthwhile to first identify the extreme events in a market and then investigate the bidding behaviour of generators around these events. In this regard, an important feature of deregulated electricity markets world-wide is the intermittent occurrence of abnormally high prices or price spikes in the spot electricity market (Barlow, 2002; Escribano et al., 2002; Lucia and Schwartz, 2002; de Jong and Huisman, 2003; Byström, 2005; Cartea and Figueroa, 2005). Both the size of these irregular price events and their duration are particularly harmful to electricity retailers who cannot pass on price risk to customers (Anderson et al., 2007). Irregular price events (or price spikes) occur when the price of electricity exceeds a given price threshold. In the Australian market, the dispatch price fluctuates between A$30 and A$70/MWh in normal conditions and the threshold of spike that is often used in the literature is usually in the region of A$80 - 100/MWh (Becker et al., 2007; Christensen et al., 2009, 2012; Clements et al., 2013).

The threshold for spike is set at a certain level for that normally represents the cost of generation given the current generation technology. However, in the NEM, the dispatch price sometimes can reach the market floor (A$-1,000/MWh) and cap (A$13,500/MWh) prices set by AEMO. These extreme price events have obviously failed to be representative for the

[2]For more details, see, Australian Energy Market Commission (2015)

underlying cost of generation. And the occurrence of these extreme dispatch prices has two significant effects on the market: First, according to the price setting rules, the trading price in a half-hourly interval will be strongly affected by an extreme dispatch price in that half-hour interval due to the averaging of the six dispatch prices. Second, despite that the extreme dispatch price might be a nodal one that is only caused by a mismatch of the equilibrium in a small area, it directly raises the trading price for the whole region as the regional trading price is taken as the highest nodal price in the region. Consequently, these extreme events may have impact on all the generators in the regional market. Moreover, according to the flexible bidding rules that allow generator to change bids five minutes before dispatch, it then becomes important to ask what would generator response to these extreme price variations in terms of their bidding behaviour, and whether the occurrence of these extreme price events and the bidding behaviours of generators around these events are aligned with the intention of the market design.

To answer these questions from a regulatory point of view, Chapter 6 of this thesis is concerned with an important policy conundrum faced by regulators in electricity markets world-wide, namely, the need to ensure the reliability of the supply while simultaneously promoting market efficiency through competition (Zhang, 2009; Bosco et al., 2012; Lízal and Tashpulatov, 2014; Bustos-Salvagno, 2015). This question is addressed in the institutional context of the NEM and particularly the regional market for the state of Queensland. In July 2007, the Queensland government started to implement a move towards full retail competition in the market. The operation of the NEM in Queensland changed significantly with the sale of two partly government-owned energy retailers and further reductions in government-owned generation capacity.[3] In another structural development of note, privately-owned electricity retailers embarked on a substantial program of investment in generation capacity mainly in the form of gas-fired turbines. These plants have a higher marginal cost than the base-load (mainly coal-fired) generators and therefore are used primarily as marginal (peaking) generators.

[3] State Government involvement in electricity generation in 2014 stands at about 50% down from around 65% in 2007 and there is no government presence in retailing.

Along with these structural changes, the market rules of the NEM were largely unchanged. The most significant exception was the large increase in the level of the market price cap prior to the introduction of competition and subsequent increases since then (from A$5,000/MWh to above A$13,500/MWh). The reason for the successive changes in the price cap was to facilitate competition by giving participants more freedom to conduct their trading activities free from a binding price cap constraint. Another reason for having such a high cap price is due to the 'missing money' issue that is related to marginal generators (Cepeda and Finon, 2011; Briggs and Kleit, 2013). That is, marginal generators in a market have the essential role of preventing black-outs in peak demand periods. Nonetheless, due to the high marginal cost and the limited peak demand periods, marginal generators are normally dispatched and paid only in the limited peak demand intervals. Thus the prices at these intervals need to be high enough for allowing marginal generators to cover both their fixed and marginal costs, and consequently to be able to exist. So in part, raising the market cap price of the NEM to the current level of A$13,100/MWh reflects this necessity. The fundamental question remains, however, of whether or not increases in the market cap actually had the (unintended) consequence of promoting strategic behaviour by market participants which in turn significantly affected electricity prices.

A re-evaluation of the policy rules governing the NEM is overdue. Since the advent of deregulation almost 20 years ago, the rules of engagement have evolved and changed without being subject to rigorous examination (for studies of the NEM, see, Wolak, 2000; Outhred, 2000; Short and Swan, 2002; Hu et al., 2005; Tamaschke et al., 2005). There is certainly enough casual empirical evidence to support the claim that prices in the NEM have become increasingly volatile in recent years. The crucial question of course is whether this volatility is linked to strategic behaviour and therefore whether current policy with respect to regulation of the electricity market needs to be revisited. To address these questions, a detailed dataset based on 'DispatchIS Reports', 'Yesterdays Bids Reports' and 'Next Day Dispatch', all of which is publicly available from the market operator, is built.[4] Data for dispatch prices, dispatched quantities, loads, inter-regional flows and bids and actual dispatch

[4]See www.nemweb.com.au.

of every generator in the market are recorded at a five-minute frequency for the period from 04:05, 30 November 2008 to 04:00, 24 June 2014, a total of 681,696 observations, is used to explore the bidding behaviour of market participants.

2.4 Conclusion

In concluding this chapter, the deregulation of electricity markets world-wide has generated a large literature with regard to every aspect of the markets. However, as the deregulation process deepened over the years, the structure of the markets and the operation of all the market participants have also evolved based on the experience accumulated. Consequently, a continuous effort for deepening the understanding of the operation of the market is required for both policy makers and all other market participants in order to maintain and further the achievement of the market deregulation. Moreover, as the data about the operation of deregulated market become more readily available and detailed, it provides a unique opportunity at this point to explore the details of the operation of deregulated electricity markets especially of which in connect with the unique characteristics of electricity market.

Chapter 3

Forecasting day-ahead electricity load using a multiple equation time series approach

3.1 Introduction

The focus of this chapter is short-term pre-dispatch (up to 24 hours ahead) load forecasts for the Queensland region of the NEM, using half hourly data for the period from 12th July 1999 to 27th November 2013. The reasons for the importance of accurate short-term load forecasting differ for each of the players in the market. From the perspective of the market operator (NEM), forecasting is crucial to the scheduling and dispatch of generation capacity; for the electricity generators, the strategic choices involved in bidding and rebidding of capacity depend critically on load forecasts; and for the electricity retailers, load forecasting affects decisions about the balance between hedging and spot acquisition of electricity. For these reasons, short-term load forecasting remains a problem of central interest and one which has generated a large literature.

Statistical models for short-term load forecasting fall very naturally into three main categories. *First*, single equation time series models model the trajectory of load using traditional time series methods (Hagan and Behr, 1987; Darbellay and Slama, 2000; Taylor

and McSharry, 2007). The efficacy of this approach derives from the strong seasonal patterns in electricity load. *Second*, and probably the current method of choice for practitioners, is the neural network approach in which the trajectory of load is modelled semi-parametrically using basis functions with emphasis on the non-linearity of load (Park et al., 1991; Zhang et al., 1998; Hippert et al., 2001). *Third*, multiple equation time series models have enjoyed some popularity in the literature but their influence has waned in recent years. In this approach, each period of the day (usually each half hour or hour) is treated as a separate forecasting problem with its own equation (Peirson and Henley, 1994; Ramanathan et al., 1997; Espinoza et al., 2005; Soares and Medeiros, 2008).

The central aim of this chapter is to demonstrate that the multiple equation approach has the potential to achieve a very competitive forecast accuracy. The advantages of the approach are that the explanatory factors driving forecast performance are visible, testable using traditional tests and the fact that the model specification is linear in parameters meaning that ordinary least squares can be used to estimate the parameters rather than a numerical optimisation algorithm. The seminal paper on the multiple equation approach to load forecasting is that of Ramanathan et al. (1997) in which the advantage of the multiple equation approach was first demonstrated in the context of the Californian electricity market. In the Australian electricity market, a Bayesian approach is employed by Cottet and Smith (2003) to a multiple equation model in a case study of the regional market of New South Wales. Perhaps the most insightful multiple equation model is that of Cancelo et al. (2008) who build a model of load in the Spanish electricity market.

What distinguishes the proposed model in this chapter from its predecessors in the multiple equation time series tradition is the way in which the daily and weekly patterns in electricity load interact and also the recognition of the importance of intra-day correlation in load. It turns out that allowing for a distinct weekly pattern in the coefficients governing one-day lagged load is a crucial advance on previous work. The efficacy of this innovation in dealing with seasonality is demonstrated by comparing it with two traditional ways of dealing with seasonality, namely the double seasonal ARIMA, and the double seasonal Holt-Winters exponential smoothing approach (Gould et al., 2008). Incorporating the proposed

refinements into a multiple equation model, the forecasting performance of the final chosen model is shown by comparing with the multiple equation model of Cancelo et al. (2008) and a semi-parametric approach used by AEMO.

In Section 3.2, a prototype model representing the starting point for the modelling exercise is developed. This model includes a piecewise linear response of load to temperature and the development of load variations for special days (public holidays). In Section 3.3, the prototype model is expanded to capture detailed seasonality and intra-day dependency of load. Focusing on comparing the effectiveness of modelling the seasonality of load, Section 3.4 compares the proposed model with two other popular alternatives. Section 3.5 presents the important forecasting results. The 12-hour ahead forecast accuracy of the proposed model is compared with the forecasts from the AEMO and an alternative multiple equation model of Cancelo et al. (2008). Section 3.6 is a brief conclusion.

3.2 A prototype multiple equation model

To provide a perspective on the forecasting problem addressed in this chapter, Figure 3.1 plots the average half-hourly load over a day and average half-hourly load over the period of a week using Queensland data with the average taken over the entire sample period from 12th July 1999 to 27th November 2013. Diurnal and weekly patterns, both well documented features of electricity load (Engle et al., 1989; Harvey and Koopman, 1993; Taylor, 2010), are clearly evident. Load picks up very quickly between the hours of 06:00 and 08:00 from the overnight low and remains high during the daylight hours. The daily peak in the load profile usually occurs at 18:00 before tailing off once more. The weekly pattern in load is also quite pronounced with a regular load profile evident from Monday through Thursday, but with significant differences on Friday, Saturday and Sunday. While it is tempting to seek to model the trajectory of load making use of these well defined features, in fact this turns out to be a sub-optimal strategy. The averaging process involved in computing the quantities in Figure 3.1 smooths out much of the half-hourly variation in load and it is this variation that a good forecasting model must capture.

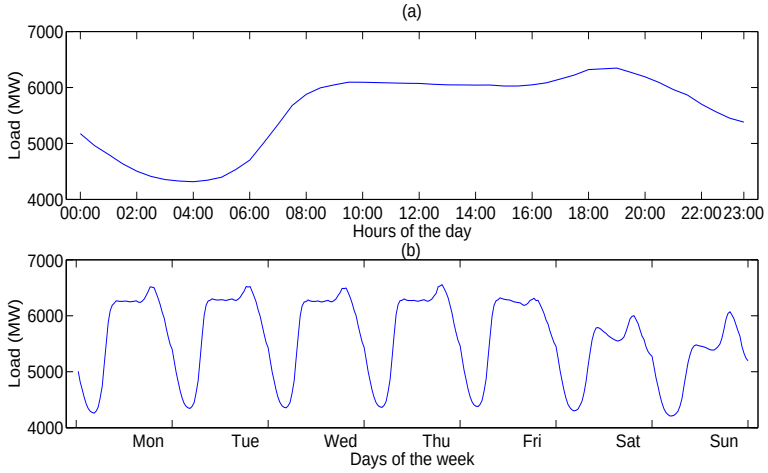

Fig. 3.1 Averaged half-hourly load over a day and averaged half-hourly load over a week in Panels (a) and (b) respectively, for Queensland over the period from 12th July 1999 to 27th November 2013.

3.2.1 The basic model structure

A model structure that captures half-hourly variability in load while respecting the features of the load profile in Figure 3.1 is one in which each half hour is modelled separately, but also uses the diurnal and other seasonal information in the load series. Let the logarithm of the load at half hour h and day d be given by L_{hd}, then, the ARMA structure of the prototype model for a given half hour period is

$$L_{hd} = \theta_{h0} + \theta_{h1} L_{hd-1} + \theta_{h2} L_{hd-7} + \phi_{h1} \varepsilon_{hd-1} + \phi_{h2} \varepsilon_{hd-7} + \varepsilon_{hd},$$

in which $h = 1, \cdots, 48$ and ε_{hd} is the disturbance term. So for each half-hour, h, the parameters are estimated based on a subset of the data which only contains the observations at that interval. In this way, the partial correlation between load and lagged load are allowed to differ in a daily pattern by the different parameter values across equations. A minimal lag structure requires L_{hd} to be explained by load in the same half hour on the previous day, L_{hd-1} and the load in the same half hour of the same day in the previous week, L_{hd-7}. For the same reasoning, the unexpected changes in load in the same half hour on the previous day, ε_{hd-1} and the previous week, ε_{hd-7}, are included.

It is important to factor in the effects of public holidays into the load forecasting equation, something which is accomplished quite parsimoniously using dummy variables following Cottet and Smith (2003) and Espinoza et al. (2005). To economise on the number of parameters to estimate, these special days are categorised into six distinct groups. Good Friday, Easter Monday, Christmas Day and New Year's Day are the four unique special days. The remaining two groups are a local Brisbane (the capital city of Queensland) only holiday and all the single day public holidays. Including special day variables, the prototype model becomes:

$$\begin{aligned} L_{hd} =& \theta_{h0} + \theta_{h1} L_{hd-1} + \theta_{h2} L_{hd-7} + \phi_{h1} \varepsilon_{hd-1} + \phi_{h2} \varepsilon_{hd-7} + \varepsilon_{hd} \\ &+ \sum_{j=1}^{6} \left(\alpha_{jh1} \mathbb{S}_{jhd} + \alpha_{jh2} \mathbb{S}_{jhd-1} \right) , \end{aligned}$$

where, $\mathbb{S}_{jhd}$ is the jth type of special day at half-hour interval h of day d. Following Ramanathan et al. (1997), the effect of one day lagged special days, $\mathbb{S}_{jhd-1}$, is also considered. The reasoning is that when the load on special days (which is typically lower than on a normal day) is used as one day lagged load, L_{hd-1}, to infer load on normal days, the effect can be suitably adjusted. This adjustment is found to be significant and is therefore maintained. The effect of one week lagged special holidays is also investigated but discounted because the improvement was found to be insignificant.

3.2.2 Dealing with the effect of temperature

There is some evidence to suggest that the response of load to temperature is nonlinear in nature and the challenge is to model this nonlinear response but at the same time maintain a model specification that is linear in parameters. A piecewise linear specification following Cancelo et al. (2008) is adopted with linear responses in four different temperature ranges: 9°C - 15°C, 9°C - 20°C, 22°C - 26°C and 22°C - 30°C. Temperatures between 20°C and 22°C are regarded as comfortable and having no extra effect on load. Also the temperatures beyond 9°C and 30°C are also treated as having no extra effect since the demand is ultimately limited by the capacity of temperature controlling devices, an effect termed as exhaustion. If

temperature in half hour h on day d is denoted T_{hd}, then to implement the piecewise linear specification four variables must be constructed which represent the changes in the relevant ranges of temperature. For the cooling degree temperature ranges the following two variables are defined:

$$\mathbb{C}_{1hd} = \begin{cases} 0 & T_{hd} \leq 22 \\ T_{hd} - 22 & 22 < T_{hd} \leq 30 \\ 30 - 22 & 30 < T_{hd} \end{cases}, \quad \mathbb{C}_{2hd} = \begin{cases} 0 & T_{hd} \leq 22 \\ T_{hd} - 26 & 26 < T_{hd} \leq 30 \\ 30 - 26 & 30 < T_{hd}. \end{cases}$$

Similarly, for the heating degree temperatures another two variables are defined:

$$\mathbb{H}_{1hd} = \begin{cases} 0 & 15 \leq T_{hd} \\ 15 - T_{hd} & 9 \leq T_{hd} < 15 \\ 15 - 9 & T_{hd} < 9 \end{cases}, \quad \mathbb{H}_{2hd} = \begin{cases} 0 & 20 \leq T_{hd} \\ 20 - T_{hd} & 9 \leq T_{hd} < 20 \\ 20 - 9 & T_{hd} < 9 \end{cases}.$$

These variables together admit a piecewise linear response of load to temperature as illustrated in Figure 3.2 which is similar in spirit to the flexible spline method used by Harvey and Koopman (1993). The ranges of defined temperature variables in which they takes non-zero values are denoted by solid lines with arrows indicating the direction of the values which deviate positively from zero. Also shown is a nonparametric kernel regression of the conditional expectation of load given temperature. The nonlinear nature of the relationship is apparent, but the piecewise linear fit appears almost identical to the nonparametric regression. The advantage of the piecewise linear specification is that it accommodates the nonlinearity but does so within a model that remains linear in parameters. It should be noted that different combinations of knots for specifying temperature variable were tried in the final version of the model, but discarded in favour of the current specification. Although, the temperature variables are included in the model, the actual load plot in Figure 3.2 suggests that load varies quite widely for any given temperature. This may be a consequence of the diverse

climate in Queensland and the non-representative temperature record which is taken at only one specific location.

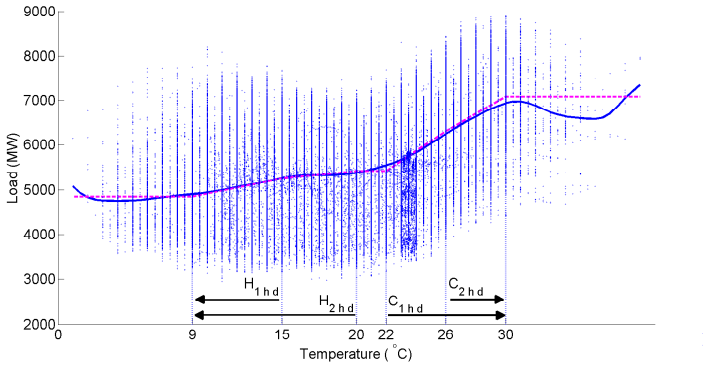

Fig. 3.2 Queensland load and temperatures, from July 1999 to December 2013. Solid line denotes a nonparametric regression fit with normal kernel and bandwidth 1. Dashed line is the ordinary least squares fit with the four temperature variables $\mathbb{C}_{1hd}$, $\mathbb{C}_{2hd}$, $\mathbb{H}_{1hd}$ and $\mathbb{H}_{2hd}$. The ranges of defined temperature variables in which they deviate positively from zero are indicated by the arrows.

Incorporating the temperature variables into the prototype model yields

$$\begin{aligned} L_{hd} =& \theta_{h0} + \theta_{h1} L_{hd-1} + \theta_{h2} L_{hd-7} + \phi_{h1} \varepsilon_{hd-1} + \phi_{h2} \varepsilon_{hd-7} + \varepsilon_{hd} \\ &+ \sum_{j=1}^{6} \left(\alpha_{jh1} \mathbb{S}_{jhd} + \alpha_{jh2} \mathbb{S}_{jhd-1} \right) \\ &+ \sum_{k=1}^{2} \left(\beta_{kh1} \mathbb{H}_{khd} + \beta_{kh2} \mathbb{C}_{khd} + \beta_{kh3} \mathbb{H}_{khd-1} + \beta_{kh4} \mathbb{C}_{khd-1} \right). \end{aligned} \tag{3.1}$$

This is the preferred specification for the prototype model against which all the refinements in later sections will be judged.

3.2.3 Estimating and forecasting the prototype model

The prototype model in (3.1) can be estimated equation-by-equation using iterative ordinary least squares (Spliid, 1983). In the estimation, each equation is initially estimated ignoring the moving-average error terms and the regression residuals stored. The equations are then re-estimated using the regression residuals from the previous step as observed moving average

error terms. This process is then iterated until convergence which is defined as the difference in parameter values in successive iterations being less than a user supplied tolerance, in this case the square root of machine precision for floating-point arithmetic.

To assess forecast performance, a 3-year rolling window of data is used for model estimation. The day-ahead forecast is produced starting from 00:00 and uses the information available at the time of making the forecast with the exception of the temperature variables. To avoid having to provide forecasts for temperature, the actual data are used in all forecasting evaluations unless specified otherwise. Moreover, as the next-day temperature forecasts are very accurate in general, any loss in accuracy of load forecast is expected to be very small when the actual temperature is replaced with a forecast. The models are re-estimated every week. In total, a period of over 11 years from July 2002 to December 2013 is used for forecast evaluation. MAPE is used as the main criterion for assessing forecast accuracy.

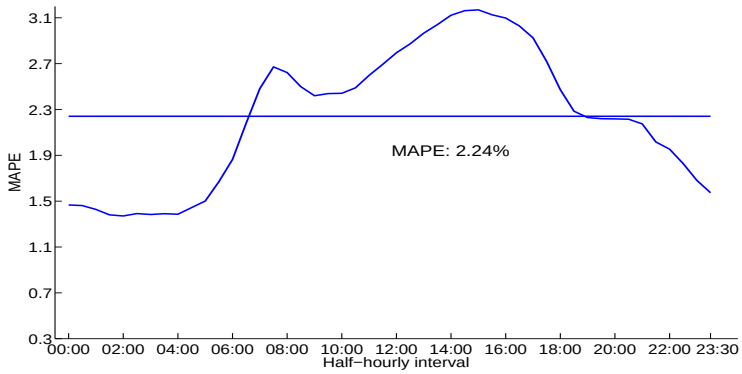

Fig. 3.3 Half-hourly MAPEs and overall MAPE for the prototype model, Equation (3.1). The overall MAPE is denoted as the solid horizontal line with its value indicated below.

A summary of the forecasting results for the prototype model are reported in Figure 3.3. The overall MAPE obtained is 2.24% with half-hourly MAPEs during the daily peak period slightly over 3%. Figure 3.3 also shows a clear daily pattern in half-hourly MAPE in which it reaches its lowest point during the night hours, increases to a small peak at around 08:00 and then rises continually to the daily maximum at around 16:00.

3.3 Extensions to the prototype model

The importance of seasonal patterns in load for accurate load forecasting is apparent and well documented in the literature (Engle et al., 1989; Harvey and Koopman, 1993; Taylor, 2010). In this section two extensions to the prototype model in (3.1) are proposed. The first extension addresses the important interaction between daily and weekly load patterns, and the second deals with intra-day load dependency by treating the equations as a recursive system.

3.3.1 Addressing seasonality

Although the design of the lag structure in Equation (3.1) is based on observed load profile, it does not capture completely its seasonal features. Figure 3.4 plots the weekly pattern in the forecast errors from the prototype model in (3.1), computed by averaging the half-hourly forecasting errors over a week. It is particularly evident that load in the half-hour intervals on Saturday and Sunday is significantly over-predicted (negative bias in the errors). This stems from the fact that the generally higher load on a weekday is being used as one-day lagged load in generating the forecast for weekends. Similarly, when Sunday load is used in generating the forecast for Monday, significant under-prediction occurs (positive bias in the errors). Essentially this bias is due to the fact that the coefficients on one-day lagged load do not differentiate between days of the week. A simple way to deal with this issue is to interact the one-day lagged load with day-of-the-week dummy variables, $\mathbb{W}_{dp}$, $p = 1, \cdots, 7$. Attempts to reduce the number of dummy variables in the specification, for example by using one for weekdays and one for weekends, or defining the dummy variables in terms of whether the day before and after is a weekday or in the weekend, produced inferior results.

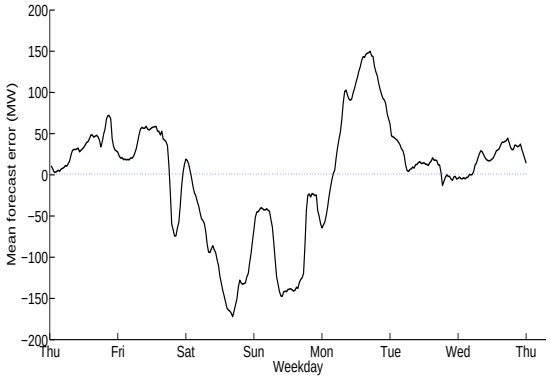

Fig. 3.4 The mean half-hourly forecast errors in days of a week over all weeks for the prototype model, Equation (3.1).

It is also possible that there is an annual pattern in the load, despite the sub-tropical nature of the Queensland climate. Similar to the treatment of the weekly pattern where the effect is channelled via the coefficient on L_{hd-1}, the annual pattern is specified in such a way that it enters the model via the coefficient on L_{hd-7}. Accordingly Fourier polynomials with annual cycles are interacted with the one-week lagged load, L_{hd-7}. The degree of the Fourier polynomials in the series expansion is four. While this choice is not tested formally, experimentation showed that little is to be gained by increasing the degree of the polynomials.[1]

Incorporating the adjustments for the weekly and annual cycles gives the extended model

$$\begin{aligned} L_{hd} =& \theta_{h0} + \theta_{hd1} L_{hd-1} + \theta_{hd2} L_{hd-7} + \phi_{h1} \varepsilon_{hd-1} + \phi_{h2} \varepsilon_{hd-7} + \varepsilon_{hd} \\ &+ \sum_{j=1}^{6} \left(\alpha_{jh1} \mathbb{S}_{jhd} + \alpha_{jh2} \mathbb{S}_{jhd-1} \right) \\ &+ \sum_{k=1}^{2} \left(\beta_{kh1} \mathbb{H}_{khd} + \beta_{kh2} \mathbb{C}_{khd} + \beta_{kh3} \mathbb{H}_{khd-1} + \beta_{kh4} \mathbb{C}_{khd-1} \right), \end{aligned} \tag{3.2}$$

[1] In principle, the weekly pattern previously discussed can also be modelled using Fourier polynomials. The dummy variable specification is preferred because it allows a natural interpretation of the coefficient estimates.

in which:

$$\theta_{hd1} = \sum_{p=1}^{7} \eta_{hp} \mathbb{W}_{dp},$$

$$\theta_{hd2} = \tau_{h1} + \sum_{q=1}^{4} \left[\tau_{h2q} \sin\left(2q\pi\left(\frac{hd}{17472}\right)\right) + \tau_{h3q} \cos\left(2q\pi\left(\frac{hd}{17472}\right)\right)\right].$$

Forecasts obtained from this model are now evaluated using exactly the same procedure as outlined in Section 3.2.3.

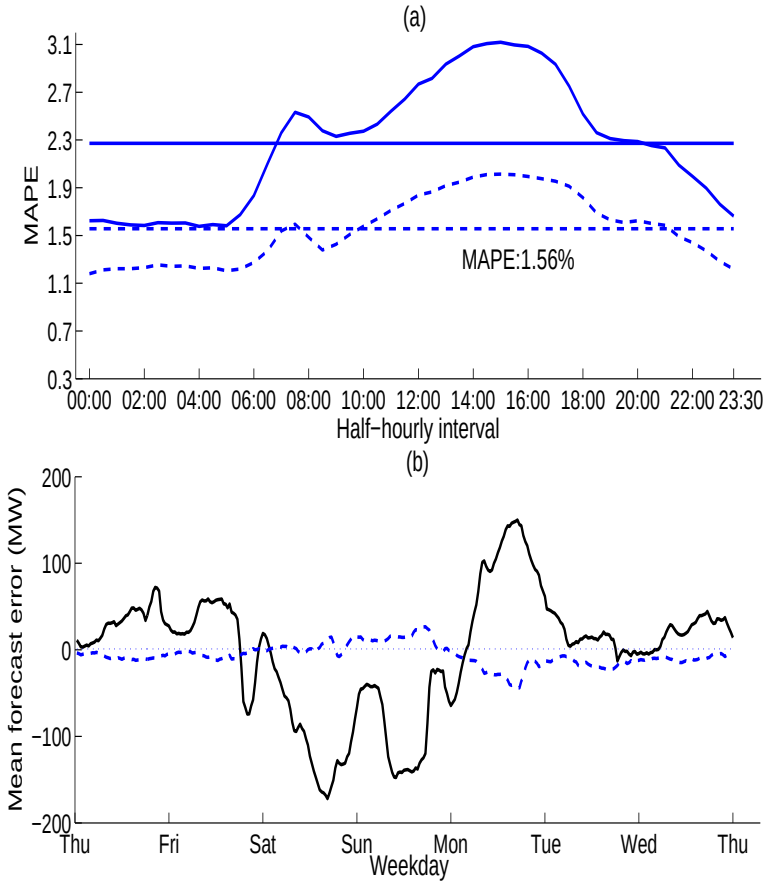

Fig. 3.5 In Panel (a), the half-hourly MAPEs and overall MAPE for the prototype model (solid lines) in (3.1) are compared to the model with seasonal patterns (dashed lines) in the parameters given in (3.2). The overall MAPEs are shown as horizontal lines with the value for Equation (3.2) indicated below. In Panel (b), the mean half-hourly forecast errors in days of a week over all weeks from the prototype model (Equation (3.1), solid line) and the model with seasonal patterns in the parameters (Equation (3.2), dashed line).

The half-hourly MAPEs are shown in Figure 3.5 together with the MAPEs of the prototype model. The extended model shows a significant improvement over the prototype model in every half hour period and for the overall MAPE recorded (1.56% versus 2.24%). Interestingly, it appears to be the weekly pattern rather than the annual cycle which drives this improvement. An overall MAPE of 1.61% was obtained from an alternative model with only specifying the weekly interactive dummy variables. The mean half-hourly forecast errors in days of a week obtained from the model in (3.2) are shown in Panel (b) of Figure 3.5. The weekly pattern in the forecast errors has been largely eliminated.

Table 3.1 The forecast comparison between the prototype model in (3.1) and the model with seasonal patterns in the parameters (Equation (3.2)).

		Overall MAPE	Maximum APE	No. APE $\geq 5\%$	No. APE $\geq 10\%$	No. APE $\geq 15\%$	No. APE $\geq 25\%$	Obs.
Overall	Eq. (3.1)	2.24%	33.58%	19702	2630	430	33	199584
	Eq. (3.2)	1.56%	26.86%	6303	542	107	4	
Normal days	Eq. (3.1)	2.12%	33.58%	13629	1940	322	16	137232
	Eq. (3.2)	1.50%	24.68%	3671	266	33	0	
Weekend	Eq. (3.1)	2.48%	24.75%	4912	517	61	0	57024
	Eq. (3.2)	1.62%	21.83%	2056	156	15	0	
Special days	Eq. (3.1)	2.89%	32.31%	1390	243	70	17	6384
	Eq. (3.2)	2.54%	26.86%	798	186	70	4	

A more detailed breakdown of the forecast performance is provided in Table 3.1. It is apparent that the most significant improvements achieved using the extended model in (3.2) are found in the forecasts on normal days and weekends. The total number of large forecast errors defined as an absolute percentage error (APE) greater than 5% is reduced by more than 10,000 instances (a 68% improvement). Overall, by interacting L_{hd-1} and L_{hd-7} with the weekly dummy variables and annual cycles, respectively, the overall MAPE of the forecast improves by 0.68% in comparison with the prototype model in Section 3.2.

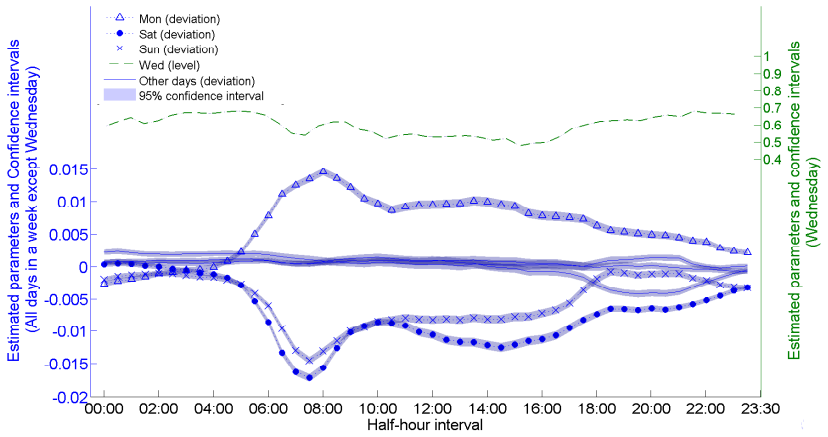

Fig. 3.6 Estimated parameters and 95% confidence intervals (shaded areas) for weekly dummy variables in the model with seasonal patterns in the parameter (Equation (3.2)). On the left vertical axis, the deflections of the parameter estimates of Monday, Saturday, Sunday and other weekdays from Wednesday are denoted by dotted line with dots, dotted line with circles, dotted line with squares and solid lines respectively. On the right axis, the level of parameter estimates for Wednesday are plotted in dashed line.

A set of representative parameter estimates for the interactive dummy variables $\mathbb{W}_{d\,p}$, $p = 1,\ldots,7$ and their 95% confidence intervals from a 3-year rolling window estimation are plotted in Figure 3.6. The largest coefficient values are seen to occur on Monday because the weekend load being used as one day lagged load in forecasting weekday load is substantially lower that the observed Monday load. The smallest coefficient values are found on the weekends. This is exactly the opposite effect to that noted for Monday; higher weekday loads are now being used to generate forecasts of lower weekend loads. More interestingly, the values of the coefficients vary in different half hours of a day. Another discernible pattern is to be found in the coefficients for different weekdays. In off-peak half-hourly intervals, the coefficients have a very similar magnitude with, in some instances, overlapping confidence intervals across different weekdays. During peak load half-hourly intervals, however, the values of the coefficients are substantially different across different weekdays. This indicates clearly that there is an interaction between daily and weekly patterns in load, a characteristic which tends to be ignored in the load forecasting literature.

3.3.2 Intra-day correlations

In the models studied thus far, the information set is defined at a daily resolution at day $d-1$. One important piece of information which is ignored is the observed load in the last half-hour period of the day prior to making a forecast, $L_{48\,d-1}$. This is particularly important for the first half hour period of the forecast, as this lagged load is observed in the immediately preceding half hour. Making this adjustment yields the model

$$\begin{aligned} L_{hd} =& \theta_{h0} + \theta_{hd1} L_{hd-1} + \theta_{hd2} L_{hd-7} + \theta_{h4} L_{48\,d-1} \\ &+ \phi_{h1} \varepsilon_{hd-1} + \phi_{h2} \varepsilon_{hd-7} + \varepsilon_{hd} \\ &+ \sum_{j=1}^{6} \left(\alpha_{jh1} \mathbb{S}_{jhd} + \alpha_{jh2} \mathbb{S}_{jhd-1} \right) \\ &+ \sum_{k=1}^{2} \left(\beta_{kh1} \mathbb{H}_{khd} + \beta_{kh2} \mathbb{C}_{khd} + \beta_{kh3} \mathbb{H}_{khd-1} + \beta_{kh4} \mathbb{C}_{khd-1} \right), \end{aligned} \tag{3.3}$$

in which

$$\begin{aligned} \theta_{hd1} &= \sum_{p=1}^{7} \eta_{hp} \mathbb{W}_{dp}, \\ \theta_{hd2} &= \tau_{h1} + \sum_{q=1}^{4} \left[\tau_{h2q} \sin\left(2q\pi \left(\frac{hd}{17472} \right) \right) + \tau_{h,3,q} \cos\left(2q\pi \left(\frac{hd}{17472} \right) \right) \right]. \end{aligned}$$

Figure 3.7 compares the MAPE of the model in (3.3) with the model (3.2) in Section 3.3.1. Not surprisingly, the biggest improvement is found in the first half-hour interval. Moreover, the substantial improvements in the half-hourly MAPEs in the first 20 half-hour intervals indicates that this idea is well worth pursuing a little further. Indeed, it is reasonable to posit that the load in consecutive half hours will be correlated so that in addition to observed load in the last half-hour period of the day prior to making a forecast, $L_{48\,d-1}$, each equation contains the lagged load from the immediately preceding half hour, $L_{h-1\,d}$. Additional lags of consecutive half-hour periods were tried, but the improvement in forecast performance was minimal.

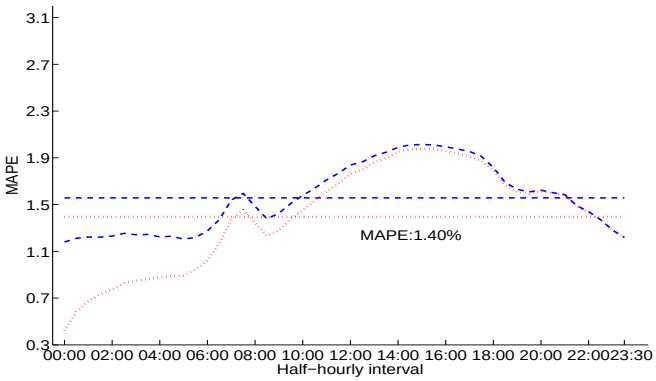

Fig. 3.7 The half-hourly MAPEs and overall MAPE for the mode with seasonal pattern in the parameters (Equation (3.2), dashed lines) and the model with the most recent load information in (3.3) (dotted lines). The overall MAPEs are shown as horizontal lines with the value of which for Equation (3.3) indicated below.

Consequently, the preferred multiple equation time series model is now

$$\begin{aligned} L_{hd} =& \theta_{h0} + \theta_{hd1} L_{hd-1} + \theta_{hd2} L_{hd-7} + \theta_{h4} L_{48d-1} + \theta_{h5} I_{h>1} L_{h-1d} \\ &+ \phi_{h1} \varepsilon_{hd-1} + \phi_{h2} \varepsilon_{hd-7} + \varepsilon_{hd} \\ &+ \sum_{j=1}^{6} \left(\alpha_{jh1} \mathbb{S}_{jhd} + \alpha_{jh2} \mathbb{S}_{jhd-1} \right) \\ &+ \sum_{k=1}^{2} \left(\beta_{kh1} \mathbb{H}_{khd} + \beta_{kh2} \mathbb{C}_{khd} + \beta_{kh3} \mathbb{H}_{khd-1} + \beta_{kh4} \mathbb{C}_{khd-1} \right), \end{aligned} \tag{3.4}$$

in which

$$\begin{aligned} \theta_{hd1} &= \sum_{p=1}^{7} \eta_{hp} \mathbb{W}_{dp}, \\ \theta_{hd2} &= \tau_{h1} + \sum_{q=1}^{4} \left[\tau_{h2q} \sin\left(2q\pi \left(\frac{hd}{17472} \right) \right) + \tau_{h3q} \cos\left(2q\pi \left(\frac{hd}{17472} \right) \right) \right], \end{aligned}$$

and $I_{h>1}$ denotes an indicator function which is equal to 1 when $h > 1$ and 0 otherwise. This modification turns the 48 equations for the half hours of a day into a recursive system. Once again, repeated application of ordinary least squares can be used to estimate the system, it provides a parsimonious way of capturing the intra-day load correlation without increasing

computational complexity significantly. Experimentation indicates that the more efficient estimation method with taking into account intra-day error correlation does not generally improve forecast accuracy.

The forecast results using (3.4) are plotted in Figure 3.8. Overall, the results show that half-hourly day-ahead MAPEs are all below 2%, with an overall MAPE of 1.36%. The improvement in forecast accuracy from using the recursive system is mainly for the daily peak intervals between 14:00 and 18:00.

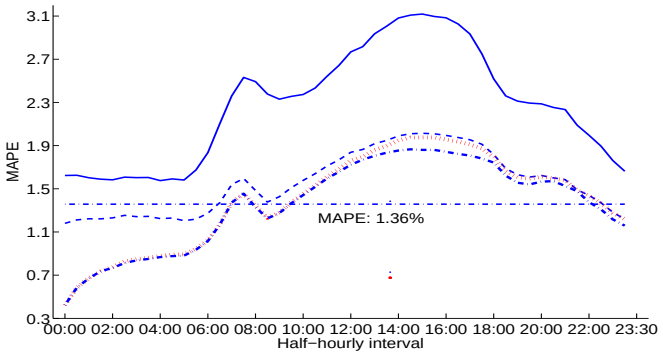

Fig. 3.8 Forecast comparison on the half-hourly MAPEs for all the four models (solid line for Equation (3.1), dashed line for Equation (3.2), dotted line for Equation (3.3) and dash-dot line for Equation (3.4)) studied and the overall MAPE for the model using recursive system in (3.4) (dash-dot horizontal line with its value indicated below).

More detailed results are reported in Table 3.2, in which models from (3.1) to (3.4) are compared. It can be seen that the most significant improvement is obtained due to the introduction of the weekly dummy variables interacting with the lagged load, L_{hd-1}, in Equation (3.2). In particular, the number of instances of large errors (APE $\geq 5\%$) decreases by nearly 10,000 on normal days when moving from the specification in the prototype model (3.1) to the weekly dummy variable specification in (3.2). In addition, incorporating the most recent information, Equation (3.3), and using a recursive system for intra-day correlation, Equation (3.4), also improve accuracy but the size of the improvement is not as large. Overall, comparing the final model in (3.4) with the prototype model in (3.1), the reduction in the number of large APE is over 70% in all bands, and overall MAPE drops from 2.24% to 1.36%, results which vindicate the modifications proposed in this section.

Table 3.2 The forecasting accuracy for the models studied, from Equations (3.1) to (3.4).

		Overall MAPE	Maximum APE	No. APE $\geq 5\%$	No. APE $\geq 10\%$	No. APE $\geq 15\%$	No. APE $\geq 25\%$	Obs.
Overall	Eq. (3.1)	2.24%	33.58%	19702	2630	430	33	199584
	Eq. (3.2)	1.56%	26.86%	6303	542	107	4	
	Eq. (3.3)	1.40%	25.98%	5130	467	93	4	
	Eq. (3.4)	1.36%	25.70%	4499	451	95	1	
Normal days	Eq. (3.1)	2.12%	33.58%	13629	1940	322	16	137232
	Eq. (3.2)	1.50%	24.68%	3671	266	33	0	
	Eq. (3.3)	1.35%	24.00%	2982	216	30	0	
	Eq. (3.4)	1.31%	21.62%	2544	203	31	0	
Weekend	Eq. (3.1)	2.48%	24.75%	4912	517	61	0	57024
	Eq. (3.2)	1.62%	21.83%	2056	156	15	0	
	Eq. (3.3)	1.44%	21.55%	1663	143	10	0	
	Eq. (3.4)	1.41%	21.77%	1521	142	13	0	
Special days	Eq. (3.1)	2.89%	32.31%	1390	243	70	17	6384
	Eq. (3.2)	2.54%	26.86%	798	186	70	4	
	Eq. (3.3)	2.28%	25.98%	697	168	59	4	
	Eq. (3.4)	2.25%	25.70%	641	167	58	1	

3.4 A comparison of approaches to modelling seasonality

The extensions proposed in Section 3.3 are designed to effectively model the detailed seasonality in electricity load. In this section, the extended model in Equation (3.4) is compared with two popular methods commonly used in the literature for dealing with seasonality. In order to focus the comparison on modelling seasonality alone, the models in this section use only lagged load information and all other information, such as temperature and special days, are ignored. The two models used for comparative purposes are now outlined.

Single equation double seasonal ARIMA model

ARIMA type models for load forecasting are widely used in the literature (Taylor, 2012; Kim, 2013). The single equation double seasonal ARIMA model is specified as:

$$\phi_p(B)\Phi_{P_1}(B^{S_1})\Phi_{P_2}(B^{S_2})(1-B)^d(1-B^{S_1})^{D_1}(1-B^{S_2})^{D_2}(L_t-c-bt) = \theta_q(B)\Theta_{Q_1}(B^{S_1})\Theta_{Q_2}(B^{S_2})\varepsilon_t, \quad (3.5)$$

where, B is the back shift operator. $\phi_p(B)$, $\Phi_{P_2}(B^{S_1})$, $\Phi_{P_2}(B^{S_2})$ and $\theta_q(B)$, $\Theta_{Q_2}(B^{S_1})$, $\Theta_{Q_2}(B^{S_2})$ denote the autoregressive and moving average parts respectively, with back shift polynomials of degree p, P_1, P_2 and q, Q_1, Q_2 respectively and seasonal factors S_1 and S_2. D_1, D_2 are the orders of differencing. The parameter c is the constant term and b is the parameter for the time trend t. The model can be written as

$$ARIMA(p,q,d)\times(P_1,Q_1,D_1)_{S_1}\times(P_2,Q_2,D_2)_{S_2}\,.$$

Focusing on comparing the effectiveness of the models for modelling the seasonality and to make the model in (3.5) comparable in a sense that it uses approximately the same amount of information as used by the multiple equation model, the specification

$$ARIMA(1,1,0)\times(1,1,1)_{48}\times(1,1,1)_{336}\,,$$

is chosen. Depending on specific case, both the proposed multiple equation model and the single equation double seasonal ARIMA in (3.5) can be easily expanded to accommodate more distant lags and other explanatory variables.

Double seasonal Holt-Winters exponential smoothing model

In short-term load forecasting, the seasonal Holt-Winters exponential smoothing (HWES) is another common choice for modelling seasonality in load (Gould et al., 2008; De Livera et al., 2011; Taylor, 2012). An intra-day cycles double seasonal HWES approach of Gould et al. (2008) is implemented here, which includes an unconstrained seasonal updating scheme with 7 daily sub-cycles in a week and additive seasonal components. As suggested by Taylor

(2012), an AR(1) term for the residual is included for better forecast accuracy. The model is specified as:

$$\begin{aligned} L_t &= l_{t-1} + b_{t-1} + \mathbf{x}_t' \mathbf{s}_{t-48} + \phi r_{t-1} + \varepsilon_t, \\ r_t &= L_t - l_{t-1} - b_{t-1} - \mathbf{x}_t' \mathbf{s}_{t-48}, \\ l_t &= l_{t-1} + b_{t-1} + \alpha r_t, \\ b_t &= b_{t-1} + \beta r_t, \\ \mathbf{s}_t &= \mathbf{s}_{t-48} + \Gamma \mathbf{x}_t r_t, \end{aligned} \tag{3.6}$$

where l_t and b_t are the level and trend at time t, respectively. The variable $\mathbf{x}_t$ is a 7×1 vector of day-of-the-week dummy variables, $\mathbf{s}_t$ is a 7×1 vector of seasonal components for the same half-hour intervals for the 7 days in a week, r_t and ε_t are, respectively, the residual term and an independent and identically distributed error term with zero mean. The constants α, β are smoothing parameters for the level and the trend, respectively, and ϕ is the AR(1) parameter for the residual. The matrix Γ has dimension 7×7 and contains the smoothing parameters for the seasonal components.

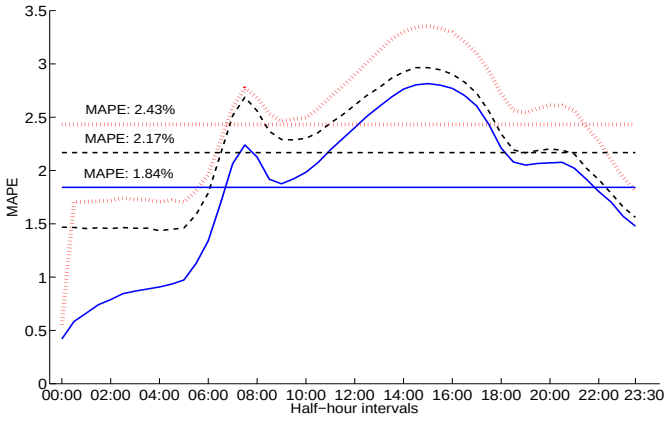

Fig. 3.9 The half-hourly MAPEs of the one day ahead forecast produced by the proposed method (Equation (3.4) without temperature and special days, denoted by solid lines), the single equation double seasonal ARIMA (Equation (3.5), denoted by dashed lines), and the unconstrained intra-day cycles double seasonal HWES (Equation (3.6), denoted by dotted lines) from July 2002 to December 2013. The overall MAPEs are shown as the horizontal lines with the values indicated above.

Figure 3.9 plots the half-hourly MAPEs of the one-day-ahead forecasts produced by the three models for the period from July 2002 to December 2013. The efficacy of the proposed multiple equation model for modelling seasonality in the load is obvious. The half-hourly MAPEs and overall MAPE for this approach are clearly lower than the corresponding forecasting statistics produced by the two competitor approaches. In short, the proposed methodology is flexible in accommodating not only daily and weekly patterns of load, but also the interaction between the two in a way that leads to a significantly improved accuracy in forecast performance as shown in Figure 3.9. In the double seasonal ARIMA, neither daily nor weekly patterns are allowed in the parameter for lagged load. In the double seasonal HWES, the unconstrained seasonal component smoothing parameters, Γ allow the seasonal component for a half-hour interval in a day of a week to be updated based on the observed load at the same half-hour interval in other days of a week, but the intra-day smoothing parameter is assumed to be fixed.

3.5 Assessing forecast performance of the full model

In this section, the forecast performance of the preferred model in (3.4) is compared against the industry standard reported by the market operator AEMO. AEMO, as the operator of the NEM, provides short-term load forecasts in pre-dispatch IS reports for the next trading day.[2] Among the horizons of the load forecast, 12-hour ahead forecasts provide important information for dispatch planning for the next day. To monitor 12-hour ahead load forecast accuracy, the monthly averaged MAPE of the 12-hour ahead forecasts is reported by AEMO as a benchmark for assessing the forecasting performance.[3] Although the details of the specification of the AEMO forecasting procedure are not available, it is known to be based on the semi-parametric specification of Fan and Hyndman (2012) and as the main forecasting

[2] See, http://www.nemweb.com.au/REPORTS/CURRENT/PreDispatchIS_Reports/.

[3] See,http://www.aemo.com.au/Electricity/Data/PreDispatch-Demand-Forecasting-Performance

model chosen by the market operator, may be taken to be representative of the state of art performance of load forecasting models.[4]

The model is also compared with the multiple equation model proposed by Cancelo et al. (2008), hereafter CEG. In this model, the seasonality of load is dealt with using a seasonal ARIMA process, which results in a non-linear model specification requiring estimation by maximum likelihood. Forecasting of the models is implemented using an identical procedure and the same set of variables defined in Section 3.2.3. To align with the 12-hour ahead forecast accuracy reported by AEMO, the accuracy of the proposed model (3.4) and CEG are assessed using 12-hour ahead forecasts.

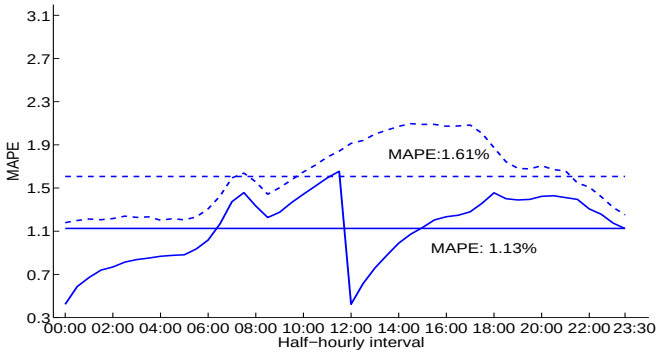

Fig. 3.10 The half-hourly MAPEs of 12-hour ahead forecast by Equation (3.4) (solid lines) and CEG (dashed lines) from July 2002 to December 2013. The overall MAPEs are shown as the horizontal lines.

A first comparison involves only the preferred model, (3.4), and CEG given that the AEMO forecast errors are only available for a shorter period. Forecasts of the two multiple equation models are generated using the same procedure as in Section 3.2.3 and the results are illustrated in Figure 3.10. It can been seen that the forecast accuracy of proposed model (3.4) is superior to that of CEG. An important anomaly in the CEG model is that it only utilises information available 24 hours previously in making a forecast. This is clearly a flaw because it does not allow the model to be flexible in terms of forecasting for periods less than 24 hours. Even in the first 12 hours when forecasts from the two models are based

[4]See, http://www.aemo.com.au/Electricity/Planning/Forecasting/National-Electricity-Forecasting-Report-2012

on the same available information, the lower MAPEs obtained from proposed model (3.4) shows the advantages of using the latest observed load together with the recursive structure developed in Section 3.3.2. Note that in the case of 12-hour ahead forecasts, the variable $L_{48\,d-1}$ in (3.4) is replaced with $\mathbb{L}_{hd} = I_{h\leq 24} L_{48\,d-1} + I_{h>24} L_{24\,d}$. This is responsible for the marked decrease in half-hourly MAPEs shown in Figure 3.10 starting from 12:00 when the most recent load information is updated. A more detailed comparison of the performance of the two models is shown in columns 2 and 3 of Table 3.3 where CEG produces inferior forecasts under all criteria. Since CEG only utilise information at a daily resolution, the results shown in column 2 for the CEG forecasts over the whole period can also be compared with the 24-hour ahead forecast from model (3.4) shown in row 5 of Table 3.2. The 1.36% overall MAPE of proposed model (3.4) is 0.25% lower than the one obtained from CEG (1.61%) and similar superior performance of the former is observed in all the criteria.

Table 3.3 Summary comparison of 12-hours ahead forecast by Equation (3.4), CEG and the AEMO forecasts.

	Jul 2002 - Dec 2013		Jul 2012 - Nov 2013			
	CEG	Eq. (3.4)	CEG	Eq. (3.4)	Eq. (3.4) without temperature	AEMO forecasts
Overall MAPE	1.61%	1.13%	1.67%	1.21%	1.37%	1.88%
Max. APE	27.99%	25.68%	20.89%	20.21%	20.26%	-
No. APE $\geq$ 5%	6981	2009	1092	384	585	-
No. APE $\geq$ 10%	645	205	128	38	44	-
No. APE $\geq$ 15%	120	45	27	7	7	-
No. APE $\geq$ 25%	3	3	0	0	0	-
Max. monthly MAPE	-	-	2.99%	1.84%	2.02%	3.2%
Obs.	199584	199584	24864	24864	24864	-

Given the limited historical data publicly available from AEMO, the period from July 2012 to November 2013 is used for subsequent comparison. Although this period is only 17 months, the advantage of the proposed model is shown clearly in Figure 3.11 and columns

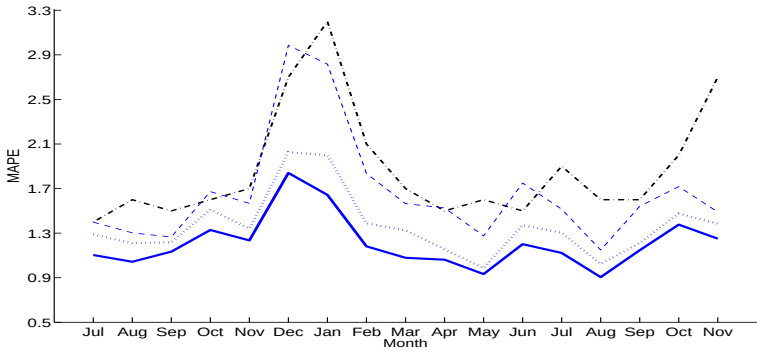

Fig. 3.11 12-hours ahead forecasts comparison of monthly MAPE between Equation (3.4) (solid line), Equation (3.4) without the future temperature (dotted line), CEG (dashed line) and the AEMO forecast (dot-dash line), from July 2012 to November 2013.

4 to 7 of Table 3.3, with the monthly MAPEs well below the AEMO forecasts and an improvement of around 0.67% in the overall MAPE over the AMEO forecasts. Since AEMO forecasts are based on temperature forecasts instead of real temperature, the results from the proposed model obtained by omitting the variables for current temperature are also reported. While there is a fall in accuracy relative to the situation when actual temperature is used, Figure 3.11 demonstrates that this effect is very small and the model is still more accurate than the AEMO forecast under all criteria (0.51% lower in the overall MAPE). The advantage of model (3.4) over CEG (which uses actual temperature data in the forecast) is also shown in Figure 3.11 and columns 4 to 6 of Table 3.3, where either with or without actual temperature, the preferred model is seen to outperform CEG under all criteria.

3.6 Conclusion

The problem of forecasting load is an important one for all electricity market participants because it informs their strategic decisions about dispatch (market operators), bidding and rebidding (generators) and trading activity (retailers). In recent times a consensus seems to have developed that neural network or non-parametric based forecasts of load, with their inherently nonlinear structure, offer the best alternative for accurate forecasting. This chapter has demonstrated that a traditional time-series approach, in which an equation is specified

for each half hour of the day, provides a viable alternative method which produces very competitive results if implemented carefully.

The multiple equation load forecasting model developed in this chapter pays particular attention to the interaction between daily and weekly load patterns. Probably the most important distinguishing factor in the proposed model relative to others in the literature is the flexibility built into the influence of load from the same half hour on the previous day. Allowing the strong weekly pattern to interact with the daily pattern in coefficients on lagged load yields important improvements in short-term forecast performance. Another innovative dimension of the current model is the use of the inherent recursive structure of the model to capture the intra-day load correlation. The effectiveness of the proposed approach on modelling the seasonal features of electricity load is demonstrated by comparing with two popular alternatives, double seasonal ARIMA and Holt-Winters exponential smoothing. Despite these modifications to the preferred model, it remains linear in parameters and can be estimated equation-by-equation by ordinary least squares.

Overall, the forecasting performance of the preferred model is impressive and significantly out-performs two benchmarks with which it is compared. In particular, the model improves on the mean average percentage error of 12-hour ahead forecast reported by the Australian energy market operator by about a third. For the entire 11 year period, the model returns a mean average percentage error of 1.36% on half-hourly day-ahead forecasts, a figure is lower than most (if not all) comparable average error statistics reported in the literature. Of course, the simple computation of an error metric does not really encapsulate the economic advantage to market participants of providing accurate load forecasts. The challenge for future work is to devise a metric that is capable of measuring economic gains to more accurate load forecasting.

Chapter 4

Forecasting quantiles of day-ahead electricity loads

4.1 Introduction

The load forecasting model proposed in Chapter 3 is useful in the decision processes of the market regulator and other market participants because it provides an accurate forecast of load that represents the future state of the demand side of the market. However, in the development of the model, or more broadly, in the load forecasting literature, the performance of the models is generally judged by measuring the difference between the forecast and the realised load. And the goal of the various proposed models is to produce a point forecast that is as close as possible to the future realised load. Admittedly, this type of point forecast can provide a very important representation for the future state of the market, especially for a rough level, but the information associated with these point forecasts is incomplete.

For example, from the perspective of a market operator, the chief goal of the dispatch planning is to meet the future demand in an economical way with adequate generation capacity to be organised in accordance with the possible future demand. If the operator only relies on a point forecast of load and prepares for dispatch accordingly, there would be around half of the time the dispatch strategy fails to meet real demand, given that load realisations follow a normal distribution conditional on the covariates included in the forecasting model.

To prevent this shortage in dispatch planning, it is natural to increase the amount of generation capacity in the planning process. Consequently, the question becomes, by how much the generation capacity should be increased so that the future demand can be met in a satisfactory way. Under-preparation of generation capacity could lead to instability of the overall system, and possibly blackouts. Whereas over-preparation results in increasing cost of generation, and loss of opportunity cost. A well-designed dispatch plan requires a balance between the security of supply and the economic viability, which could be dynamically adjusted based on the current market conditions. To achieve this balance, it is then important to not only have knowledge about an approximate level of future demand as provided by a point forecast, but also the variation of the realised demand around the forecast level. Consequently, the risk involved in relying on point forecasts and of the variation of real demand around point forecast can be fully acknowledged in the decision making process of all market participants.

In order to represent the uncertainty of future load, a relevant approach is to forecast the maximum demand in a day or a week (Engle et al., 1992; Hyndman and Fan, 2010), which is still a point forecast produced with usual mean regression models, but carries information about the upper end variation of load within a certain period. Alternatively, simulation procedures are adopted in the literature, in which simulated covariates or residuals are used to construct interval predictions for future load. An example of this type is the density forecast model of Fan and Hyndman (2012), in which, using mean regression, the variability of load forecasts arises from both simulated covariates and residuals. To pursue further on these ideas and also provide a more complete picture regarding the uncertainty of load forecast, it is natural to forecast the interval or quantiles of future load directly using quantile regressions with the quantiles being set at multiple points.

In this chapter, the possibility of using quantile regression for distributional forecast of load is explored. Although, the application of quantile regression to forecasting quantiles of various economic variables is not new, it has rarely been applied in the context of load forecasting. Instead, quantile forecasts of load are normally produced based on simulating or bootstrapping residuals or historical observations (McSharry et al., 2005; Fan and Hyndman, 2012). The reason for this phenomenon is that the classic quantile regression approach is

computationally intensive even for models with a small number of parameters, whereas in the context of load forecasting, complicated seasonal patterns, effect of temperature and special days normally require a large number of parameters to be specified to achieve a satisfactory forecasting accuracy. To overcome this computational issue while taking into account useful but vast number of covariates, López Cabrera and Schulz (2014) demonstrated using dimension reduction with principal components in the classic quantile regression framework. However, the use of principal component may lead to small variations in important covariates being ignored, resulting in loss of forecasting accuracy. In this chapter, an alternative Bayesian approach to quantile regression is used instead to accommodate a large number of parameters. Two models are proposed; each of them differ in the type of underlying residual distributions. The forecasting results are compared with each other and also with an industry standard reported by the Australia Energy Market Operator (AEMO). Overall, the proposed model performs competitively in comparison with the AEMO forecast.

4.2 Specification for the quantiles

A good load forecasting model is one which accounts for all the factors that may have significant impact on the variation of load. As shown in Chapter 3, seasonality of load, temperature and special days effects are important for forecasting load. It is also shown in Chapter 3 that the linear specification performs competitively in terms of forecast accuracy. Consequently, let y_{hd} denotes the load level at half-hour h and day d, a similar AR specification for the quantiles of load is:

$$\begin{aligned}\mathrm{y}_{hd} =& \theta_{qh0} + \theta_{qhd1}\mathrm{y}_{hd-1} + \theta_{qh2}\mathrm{y}_{hd-7} + \theta_{qh4}\mathrm{y}_{48d-1} + \varepsilon_{hd} \\ &+ \sum_{j=1}^{6} \left(\alpha_{qjh1}\mathbb{S}_{jhd} + \alpha_{qjh2}\mathbb{S}_{jhd-1}\right) \\ &+ \sum_{k=1}^{2} \left(\beta_{qkh1}\mathbb{H}_{khd} + \beta_{qkh2}\mathbb{C}_{khd} + \beta_{qkh3}\mathbb{H}_{khd-1} + \beta_{qkh4}\mathbb{C}_{khd-1}\right), \end{aligned} \tag{4.1}$$

in which $\theta_{qhd1} = \sum_{p=1}^{7} \eta_{qhp} \mathbb{W}_{dp}$, $h = 1, \cdots, 48$ and ε_{hd} is the disturbance term. The subscript q in all parameters denotes the quantile of interest. So for each half-hour, h, the parameters are estimated based on a subset of the data that only contains the observations at that interval. In this way, the daily pattern is expected to be constant in each subset, and the partial correlation between load and lagged load are allowed to differ in intra-day pattern reflected by the different parameter values across equations. A minimal lag structure requires y_{hd} to be explained by load in the same half hour on the previous day, y_{hd-1} and the load in the same half hour of the same day in the previous week, y_{hd-7}. The special days and temperature variables, $\mathbb{S}_{jhd}$ and $\mathbb{S}_{jhd-1}$ for $j = 1, \ldots, 6$ and $\mathbb{H}_{khd}$, $\mathbb{C}_{khd}$, $\mathbb{H}_{khd-1}$ and $\mathbb{C}_{khd-1}$ for $k = 1, \ldots, 2$ are defined similarly as in 3.2.

In addition, when using this AR specification, seasonality can also exist in the parameter for one day lagged load. Intuitively speaking, if the parameter for one-day lagged load is held constant, when one day lagged load is taken as Sunday load, and used for forecasting Monday load, the normally lower Sunday load will cause an under prediction. Vice-versa, an over prediction will occur if Friday load is used to forecast Saturday load. For this reason, a good forecasting model should be able to suitably adjust the parameter for one-day lagged load depending on the days of the week. This consideration leads to the parameter θ_{qhd1} to be depending on the days of a week dummy variables, $\mathbb{W}_{dp}$ with $p = 1, \ldots, 7$. Moreover, since load profile is known to be smoothly evolving, the most recent lagged information plays a very important role in inferring future load. This is especially the case in forecasting for the first half-hour period of a day when the load in the last half-hour is observed. Thus, the most recent information, y_{48d-1} is also included.

Thus far, all the covariates included in the model specification (4.1) are observed. In a normal mean regression setting, the inclusion of lagged residuals can also prove to be useful, since it allows the forecast to be dynamically adjusted based on prediction errors in the previous intervals. In quantile regression context, a similar approach is the dynamic quantile model of Engle and Manganelli (2004), in which the lagged quantiles enter the model specification. Following this idea and including the one-day and one-week lagged

quantile terms into (4.1) yields:

$$\begin{aligned}\mathrm{y}_{hd} =& \theta_{qh0} + \theta_{qhd1}\mathrm{y}_{hd-1} + \theta_{qh2}\mathrm{y}_{hd-7} + \theta_{qh4}\mathrm{y}_{48d-1} + \theta_{qhd5} f_{hd-1}(\boldsymbol{\beta}_q) + \theta_{h6} f_{hd-7}(\boldsymbol{\beta}_q) + \varepsilon_{hd} \\ &+ \sum_{j=1}^{6} \left(\alpha_{qjh1}\mathbb{S}_{jhd} + \alpha_{qjh2}\mathbb{S}_{jhd-1}\right) \\ &+ \sum_{k=1}^{2} \left(\beta_{qkh1}\mathbb{H}_{khd} + \beta_{qkh2}\mathbb{C}_{khd} + \beta_{qkh3}\mathbb{H}_{khd-1} + \beta_{qkh4}\mathbb{C}_{khd-1}\right), \end{aligned} \tag{4.2}$$

with $\theta_{qhd1} = \sum_{p=1}^{7} \eta_{qhp}\mathbb{W}_{dp}$, $\theta_{qhd5} = \sum_{p=1}^{7} \omega_{qhp}\mathbb{W}_{dp}$, and $f_{hd-1}(\boldsymbol{\beta}_q)$ and $f_{hd-7}(\boldsymbol{\beta}_q)$ denoting the one-day and one-week lagged quantiles at the same half-hour interval, respectively. The same parameter specification is used for the one-day lagged quantile as for the one-day lagged observed load for the same reasoning of the existence of weekly pattern in the coefficient for one-day lagged terms.

Assuming that the periodic patterns are modelled adequately in this specification, ε_{hd} should not have a weekly periodic pattern like y_{hd-1} does. Consequently, it is reasonable to assume that the lagged residuals would have the same impact on future load regardless of the days of the week of the residuals. With this in mind, a simple rearrangement of (4.2) leads to the specification:

$$\begin{aligned}\mathrm{y}_{hd} =& \theta_{h0} + \theta_{hd1}\mathrm{y}_{hd-1} + \theta_{hd2}\mathrm{y}_{hd-7} + \theta_{h4}\mathrm{y}_{48d-1} + \theta_{h5}\varepsilon_{hd-1} + \theta_{h6}\varepsilon_{hd-7} + \varepsilon_{hd} \\ &+ \sum_{j=1}^{6} \left(\alpha_{jh1}\mathbb{S}_{jhd} + \alpha_{jh2}\mathbb{S}_{jhd-1}\right) \\ &+ \sum_{k=1}^{2} \left(\beta_{kh1}\mathbb{H}_{khd} + \beta_{kh2}\mathbb{C}_{khd} + \beta_{kh3}\mathbb{H}_{khd-1} + \beta_{kh4}\mathbb{C}_{khd-1}\right), \end{aligned} \tag{4.3}$$

with $\theta_{hd1} = \sum_{p=1}^{7} \eta_{hp}\mathbb{W}_{dp}$. This in effect turns the dynamic quantile specification into one that is similar to a ARMA specification in mean regression. The only difference here is that the residual terms are obtained conditional on quantile estimates of all the parameters.

The final specification in (4.3) enables a conditional mean forecast for day-ahead load similar to the one proposed in Chapter 3 when ordinary least squares is used. To produce quantile forecasts instead, quantile regression of Koenker and Bassett Jr. (1978) and Koenker

(2005) may be used. Engle and Manganelli (2004) extended this framework by allowing dynamic quantiles, that is, lagged quantiles are included in the model specifications.

However, as shown by Engle and Manganelli (2004), even when the total number of parameters in the models is around five, this type of dynamic specification leads to the objective function having many local optima. Thus starting values from a grid of around five dimensions needs to be used with some local optima finding algorithm to be run many times. Also when choosing the grid of starting values, unrealistic constraints have to be imposed even on the parameter for constant (for example, constrained in the unit interval) in order to achieve the computational tractability.

To circumvent the aforementioned problems and also allow more flexibility on the number of parameters to be specified in the models, two Bayesian alternatives to the classical quantile regression are adopted and detailed in the following sections. In the first approach, a widely used, computational simple but more restrictive method is used by specifying a parametric residuals distribution. In the second approach, the residual density function is allowed to be arbitrary with the only restriction that the density be uni-modal.

4.3 Quantile estimation with a parametric residual distribution

In the classic autoregressive quantile regression of Koenker and Xiao (2006), the quantile estimate is obtained through using some local optimum search algorithm and the pin-ball loss function. To achieve a similar quantile estimate, the Bayesian quantile regression of Yu and Moyeed (2001) specifies the error term to follow an asymmetric Lapace distribution (ALD) with the density of error term ε_d at a fixed quantile q,

$$P_{dq} = \frac{q(1-q)}{\sigma}\exp\left\{-\frac{|\varepsilon_d| + (2q-1)\varepsilon_d}{2\sigma}\right\}$$

where $d = 1, \ldots, D$ with D denotes for the total number of observations (days).[1]

This error specification allows a Bayesian approach to quantile regression with the aid of Metropolis type algorithm that permits for both linear and non-linear specification for ε as a function of parameters and data. If the specification of ε is linear such as the one in 4.1. A Gibbs sampler is also available from Kozumi and Kobayashi (2011). The main idea of this Gibbs sampler is to use a mixture representation of the asymmetric Lapace distributed random variable (Kotz et al., 2001). That is, assuming ε is an asymmetric Lapace distributed random variable, it may be expressed as

$$\varepsilon = \theta z + \tau\sqrt{z}u,$$

where $\theta = \frac{1-2q}{q(1-q)}$, $\tau = \sqrt{\frac{2}{q(1-q)}}$, and z and u are independent random variables with standard exponential and standard normal distributions, respectively.

Using this mixture presentation, (4.1) can be rewritten as

$$\mathrm{y}_d = \mathbf{x}_d'\boldsymbol{\beta}_q + \theta v_d + \tau\sqrt{\sigma v_d}u_d, \tag{4.4}$$

where, $\mathbf{x}_d$ is a column vector containing the values of all covariates at day d, $\boldsymbol{\beta}_q$ is a column vector containing all the parameters at quantile q, and $v_d = \sigma z_d$.

4.3.1 Estimation procedure

Given (4.4), assigning conjugate priors with $\boldsymbol{\beta}_q \sim \mathrm{N}\left(\boldsymbol{\beta}_{q0}, \mathbf{B}_{q0}\right)$ and $\sigma \sim \mathrm{IG}(n_0/2, s_0/2)$, where $\mathrm{N}\left(\boldsymbol{\beta}_{q0}, \mathbf{B}_{q0}\right)$ denotes a normal distribution with hyper-parameters $\boldsymbol{\beta}_{q0}$ and $\mathbf{B}_{q0}$, and $\mathrm{IG}(n_0/2, s_0/2)$ denotes an inverse gamma distribution with shape parameter $n_0/2$, scale parameter $s_0/2$, the Gibbs sampler of Kozumi and Kobayashi (2011) proceeds as follows:

[1] The half-hour index, h is omitted since the 48 half-hourly subset of data with its own equations are treated as separate regression problems in this section.

1. Given a starting value or the sample from the previous iteration for $\boldsymbol{\beta}_q$, $\boldsymbol{v} = (v_1, \ldots, v_D)$ and σ, samples $\boldsymbol{\beta}_q$ with

$$\boldsymbol{\beta}_q \,|\, \boldsymbol{v}, \sigma \sim \mathrm{N}\left(\bar{\boldsymbol{\beta}}_q, \bar{\mathbf{B}}_q\right).^2$$

   In which, $\mathrm{N}\left(\bar{\boldsymbol{\beta}}_q, \bar{\mathbf{B}}_q\right)$ is a multivariate normal distribution with mean,

$$\bar{\boldsymbol{\beta}}_q = \bar{\mathbf{B}}_q \left\{ \sum_{i=1}^{D} \left(\mathbf{x}_i \left(\mathrm{y}_i - \theta v_i\right) / \tau^2 \sigma v_i\right) + \mathbf{B}_{q0}^{-1} \boldsymbol{\beta}_{q0} \right\},$$

   and covariance matrix,

$$\bar{\mathbf{B}}_q = \left\{ \sum_{i=1}^{D} \left(\mathbf{x}_i \mathbf{x}_i' / \tau^2 \sigma v_i\right) + \mathbf{B}_{q0}^{-1} \right\}^{-1}.$$

2. Then samples v_i with

$$v_i \,|\, \boldsymbol{\beta}_q, \sigma \sim \mathrm{GIG}\left(\frac{1}{2}, \bar{\delta}_i, \bar{\gamma}\right),$$

   for $i = 1, \ldots, D$. In which, $\mathrm{GIG}(v, a, b)$ is the generalised inverse Gaussian distribution with density

$$f(x|v,a,b) = \frac{(b/a)^v}{2K_v(ab)} x^{v-1} \exp\left\{ -\frac{1}{2}\left(a^2 x^{-1} + b^2 x\right) \right\}$$

   with $x > 0$, $-\infty < v < \infty$, $a > 0$, $b > 0$, $K_v(\cdot)$ denotes for a modified Bessel function, $\bar{\delta}_i = \left\{ \left(\mathrm{y}_i - \mathbf{x}_i' \boldsymbol{\beta}_q\right)^2 / \tau^2 \sigma \right\}^{1/2}$ and $\bar{\gamma} = \left(2/\sigma + \theta^2/\tau^2\sigma\right)^{1/2}$.

3. To complete a full iteration, samples σ with

$$\sigma \,|\, \boldsymbol{\beta}_q, \boldsymbol{v} \sim \mathrm{IG}\left(\frac{\bar{n}}{2}, \frac{\bar{s}}{2}\right),$$

   where, $\bar{n} = n_0 + 3n$ and $\bar{s} = s_0 + 2\sum_{i=1}^{D} v_i + \sum_{i=1}^{D} \left(\mathrm{y}_i - \mathbf{x}_i' \beta_q - \theta v_i\right) / \tau^2 v_i$.

The advantage of this approach is that all parameters of the conditional distributions have analytical solutions. Consequently the estimation procedure is computationally simple in that

[2] The conditioning on data is assumed and simplified in the notation throughout this chapter.

it only involves random drawing from some well-known distributions and is not affected by the scaling of covariates. Since all the covariates are observed in (4.1) and the specification is linear in parameters, this Gibbs sampler can be used with specification (4.1).

Differing from the autoregressive quantile regression of (4.1), the dynamic quantile regression of Engle and Manganelli (2004) introduces lagged quantiles into the model specification. This leads to the objective function needing to be evaluated recursively. As mentioned previously, it also leads to significant increases in the computational cost and restrictions on the range of starting values. Alternatively, the Bayesian approach to dynamic quantile regression is introduced first by Gerlach et al. (2011) with a Metropolis type algorithm for calculating Value-at-Risk, in which, the computational advantage is shown, since as long as the global optimum is not separated from other local optima by very low probability regions, the chain will eventually jump out of local optima and converge to the global one. This avoids running multiple optimisations on a large grid.

However, with the dynamic specification in (4.3), a large number of parameters are included. Even if it can be estimated in a Bayesian way using a Metropolis type algorithm, it involves a high dimensional parameter search problem. This usually requires artful tuning of the proposal distribution in order to achieve a reasonable acceptance rate and satisfactory convergence rate. Also, an inappropriate scaling of any covariate could possibly lead to the loss of significant inferential power from that variate. Alternatively, to reduce the dimension of the parameter search problem, observe that given the parameters for lagged error term in (4.3), θ_5 and θ_6, the remaining parameters are all linear. Consequently, the sampling of all the parameters in (4.3) can be divided into two batches as shown in the following steps with only a smaller batch to be sampled by a Metropolis type algorithm and the other batch by the Gibbs sampler just introduced:

1. Conditioning on the starting values or the previous iteration of θ_{q5}, θ_{q6}, $\boldsymbol{\nu}$ and σ. Let all the observed covariates that correspond to the remaining parameters be denoted by a D×N matrix, $\boldsymbol{C}$, where D and N denote for the total number of observations and the number of covariates, respectively. Set the starting value for lagged residuals to be 0. Define a new dependent variable, $\boldsymbol{z}$ and new covariates $\boldsymbol{w}$ as follows: Set $z_1 = y_1$ and

$\boldsymbol{w}_1 = \boldsymbol{c}_1$, then define recursively

$$z_d = y_d - \theta_{q5} z_{d-1}, \qquad \text{for } d = 2 \ldots 7,$$
$$z_d = y_d - \theta_{q5} z_{d-1} - \theta_{q6} z_{d-7}, \qquad \text{for } d > 7$$

and

$$\boldsymbol{w}_d = \boldsymbol{c}_d - \theta_{q5} \boldsymbol{w}_{d-1}, \qquad \text{for } d = 2 \ldots 7,$$
$$\boldsymbol{w}_d = \boldsymbol{c}_d - \theta_{q5} \boldsymbol{w}_{d-1} - \theta_{q6} \boldsymbol{w}_{d-7}, \qquad \text{for } d > 7.$$

2. Use the newly defined variables to sample all the parameters for observed covariates, $\boldsymbol{C}$ as in step 1 of the previous Gibbs sampler by replacing $\mathbf{y}$ for $\boldsymbol{z}$ and $\mathbf{X}$ for $\boldsymbol{W}$.

3. Conditioning on the updated parameters from the previous step, use Metropolis type algorithm to sample the two recursive parameters θ_{q5} and θ_{q5}.

An iteration of these three steps is an iteration for $\boldsymbol{\beta}_q$, then follow the steps 2 and 3 in the previous Gibbs sampler to sample $\boldsymbol{v}$ and σ, and thus complete a full iteration. This sampling approach is particularly useful in the present case, for it allows a large amount of parameter to be specified in the model for accurately capturing a variety of well-known load patterns while maintaining computational tractability.

4.4 An alternative model with a non-parametric residual distribution

The model introduced in Sections 4.2 and 4.3, denoted hereafter as Model 1, is computationally tractable due to the convenient mixture form of ALD. However, ALD can be restrictive in a sense that the scale and shape of the density is controlled entirely by three parameters. This is in sharp contrast to the classic quantile regression with its main advantage being the robustness of the quantile estimates to outliers when the pin-ball loss function is used, only

the order instead of scale of residuals can affect the estimation results. To relax the restrictive scale and shape assumption in Model 1, the approach of using a mixture of uniform kernels with Dirichlet process piror proposed by Kottas and Krnjajić (2009) is used for the residual distribution. This model will be referred as Model 2 hereafter.

Let q be the quantile of interest, in all quantile regression, the estimates are all obtained by assigning a loss function or a likelihood function that puts q total mass of probability on negative residuals and $1-q$ total mass of probability on non-negative residuals. Following this idea, defining a kernel function,

$$k(q)(\varepsilon;\sigma_1,\sigma_2) = \frac{q}{\sigma_1}\mathbf{1}_{(-\sigma_1,0)}(\varepsilon) + \frac{q}{\sigma_2}\mathbf{1}_{[0,\sigma_2)}(\varepsilon).$$

It is clear this kernel function has one mode and assigns uniform weight $\frac{q}{\sigma_1}$ for negative residuals and $\frac{q}{\sigma_2}$ for non-negative residuals that fall within the corresponding intervals.

To allow the kernel function to approximate an arbitrary density function, the popular non-parametric infinite mixture approach of using Dirichlet process (DP) (Ferguson, 1973) is used. In this case, the two parameters of the kernel function, σ_1 and σ_2 are assumed to be sampled from random discrete probability distributions, which in turn arose from two independent DPs. Consequently, the hierarchical structure is:

$$\begin{aligned}
\mathrm{y}_d \,|\, \boldsymbol{\beta}_q, \sigma_{1d}, \sigma_{2d} &\sim k_q\left(\mathrm{y}_i - \mathbf{x}'\boldsymbol{\beta}_q; \sigma_{1d}, \sigma_{2d}\right), && d = 1,\ldots,D,\\
\sigma_{rd} \,|\, G_r &\sim G_r, && r = 1,2,\ d = 1,\ldots,D,\\
G_r \,|\, \alpha_r, d_r &\sim \mathrm{DP}\left(\alpha_r, G_{r0}\right), && r = 1,2,
\end{aligned}$$

where α_r and G_{r0} are prior precision parameters and base distributions for two independent DPs indexed by r. At the highest third level, the G_r are two realisations of discrete probability distributions from the two independent DPs. Then at the second level, mixing parameters, σ_{1d} and σ_{2d} arise independently from the realised discrete probability distributions, G_1 and G_2, respectively. At the lowest level, each observed value of y_d arise from the dth kernel with the corresponding σ_{1d} and σ_{2d}.

To complete the model specification, priors for α_r and G_{r0} need to be specified. In the present case, the prior specification follow from Kottas and Krnjajić (2009), that is a Gamma distribution for each α_r and inverse Gamma with mean d_r for each G_{r0}, then a Gamma prior distribution for each k_r. In setting the hyper-prior for these gamma distributions, the shape parameter is set to 2 and scale parameter is set to 20. Different hyper-priors are also tried as a robustness check.

The sampling procedure of Model 2 follows a standard sampling approach with DP. In the first stage, σ_{1d} and σ_{2d} for all d are updated using multinomial sampling, then G_r for $r = 1,2$ are updated using the blocked Gibbs sampler of Ishwaran and James (2001) based on the stick-breaking procedure of Sethuraman (1994). In the second stage, each α_r is updated using the approach of Escobar and West (1995), then d_r for $r = 1,2$ are updated based on their conditional distribution. In the final stage, use Metropolis algorithm to update all $\boldsymbol{\beta}_q$. This sampling approach follows a general framework of the common approach for sampling from a non-parametric DP prior (Gelman et al., 2014). A detailed description for this approach when using the specific uniform kernel function can be found in Kottas and Krnjajić (2009), with the only difference here is that in the first stage of their approach, the Pólya urn representation of DP (Blackwell and MacQueen, 1973) is used with σ_{rd} sampled one at a time from its conditional distribution.

Since it is no longer possible to obtain the distribution of $\boldsymbol{\beta}_q$ conditional on the mixture of uniform kernels, all the parameters in this model need to be sampled based on a metropolis type algorithm. Since the parameter search dimension has increased dramatically in Model 2, it requires a fine tuning of the algorithm to ensure the convergence. In the present case, a natural choice for the starting values and proposal distribution are obtained from the samples of Model 1 with the Gibbs sampler. Based on these settings, the adaptive MCMC of Haario et al. (2001) is also used, which allows appropriate adjustments of the proposal distribution when running the chains.

4.5 Forecasting and performance evaluation

Forecasting with the proposed models can be obtain by attaching an extra step at the end of each iteration in the MCMC after convergence. In this step, the $\boldsymbol{\beta}_q$ updated in each iteration is used to forecast future quantile of load. In the remainder of this chapter, all results are based on using weakly informative prior with large variance or scale for the hyper-prior parameters, 20,000 iteration for burn-in period and another 20,000 samples after burn-in for posterior inference. The convergence of the sampler is examined visually, the algorithm normally reach convergence in around 2,000 iterations for specification (4.3). The acceptance rate in the Metropolis steps for (4.3) varies from 15% to 35% depending on the half-hour interval. A moving window forecasting procedure is used based on a three-year window length for model estimation, and the model is re-estimated weekly.

In terms of performance evaluation, unlike point forecasting in which the goal is to produce forecasts as close as possible to the realised value, quantiles of load are not directly observable. Consequently, different measures of accuracy have to be used. Since the goal is to forecast the quantiles of load, the coverage ratio of the forecast defined by:

$$Covarage\,ratio = \frac{\sum_{t=1}^{T} \mathbf{1}_{(-\infty,0)}(\varepsilon_t)}{T}, \tag{4.5}$$

with $T = D \times H$, becomes a simple but useful measure. Obviously, the closer the coverage ratio of the quantile forecast to the quantile of interest, the better the performance. In addition, the mean of the pin-ball loss:

$$Loss(q,\varepsilon) = (q-1)\varepsilon \mathbf{1}_{(-\infty,0)}(\varepsilon) + q\varepsilon \mathbf{1}_{[0,+\infty)}(\varepsilon),$$

is also used, which is a more informative measure, and used widely in quantile forecasting problems as an accuracy measure (Taylor, 2007; Clements et al., 2008).

Traditionally, tests for the coverage ratio and independence of the hits can also be carried out (Christoffersen, 1998; Christoffersen and Pelletier, 2004; Engle and Manganelli, 2004). However, these tests only provide limited information about the performance of the models

by giving a 'yes' or 'no' answer. Moreover, when using composite tests, the coverage ratio and independence of hits are considered in a single test. In the present case, the 48 half-hour equations are treated separately in one-day ahead forecasts. It is therefore natural for the prediction or hit to be highly correlated between half-hours of a day, thus failing the test. This issue arises naturally due to the forecasting resolution being different from that of the estimation problem instead of being due to some issues with the model specification itself. Consequently, in all the performance evaluation in this chapter, the coverage ratio and pin-ball loss functions are used as the main accuracy criteria.

Due to the importance of the information quantile forecasts carry for the decision making of market participants in an electricity market, AEMO also provide quantile forecast at 0.1, 0.5, and 0.9 quantiles. Consequently, besides comparing between the two proposed models, AEMO forecasts are also used as an industrial standard in the following comparison. As a result, the comparison is focused at the fixed quantile of 0.1, 0.5, and 0.9.

Table 4.1 A summary of the forecast accuracy of the two proposed models and AEMO forecasts based on coverage ratio and pin-ball loss criteria. The evaluation period starts from 00:30 2014-11-24 to 00:00 2015-08-31.

	Coverage ratio × 100			Pin-ball loss			Observations
	$q = 0.1$	$q = 0.5$	$q = 0.9$	$q = 0.1$	$q = 0.5$	$q = 0.9$	
(Model 1)	8.41	43.30	85.96	22.58	49.09	23.54	
(Model 2)	14.12	43.25	78.91	24.99	49.09	26.99	13,440
AEMO	4.88	38.49	89.31	26.20	53.95	24.60	

Table 4.1 shows a summary for the forecast accuracy of the three models compared across all half-hourly intervals over the entire forecasting period from 00:30 2014-11-24 to 00:00 2015-08-31. It can be seen in general, Model 1 produces the best coverage ratio and pin-ball loss in almost all the three quantiles compared. Besides, the coverage ratio and pin-ball loss of Model 2 tends to be preferable over that of the AEMO forecast at 0.1 and 0.5 quantiles, and conversely at 0.9 quantile. The only exception where Model 1 has a less preferable results is in the coverage ratio at 0.9 quantile when comparing with the AEMO forecasts (85.86% vs 89.31%).

To provide a more detailed comparison of the coverage ratios of the three forecasts, Figure 4.1 shows the coverage ratios of the three forecasts in each half-hour interval of the day. It becomes apparent that the coverage ratio of Model 1 (solid line) is preferred over the AEMO forecast (dashed line) at the 0.9 quantile. The better coverage ratio of the AEMO forecast at the 0.9 quantile (89.31%) in Table 4.1 is due to averaging between the higher ratios in first 13 half-hour intervals and the lower ones in the intervals after. Whereas the coverage ratio of Model 1 stays closer to 0.9 than the other two alternatives in almost all the 48 half-hour intervals. Similarly at quantiles 0.1 and 0.5, Model 1 also exhibits superior performance over the other two.

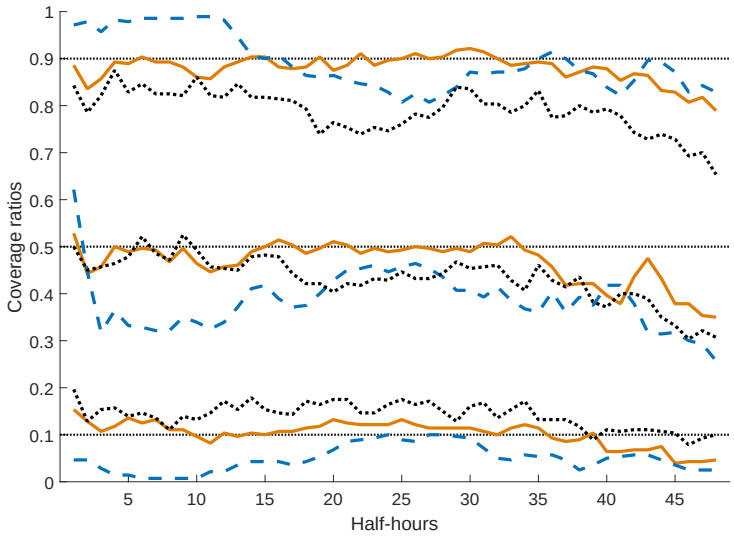

Fig. 4.1 Half-hourly coverage ratios of the three forecasts, Model 1 (solid line), Model 2 (dotted line) and AEMO forecasts (dashed line) at quantile 0.1, 0.5 and 0.9. The evaluation period starts from 00:30 2014-11-24 to 00:00 2015-08-31.

In Figure 4.2, a similar half-hourly breakdown of pin-ball losses of the three forecasts are shown. It can be seen that the losses of Models 1 and 2 stay lower than the AEMO forecast in around the first 18 and last 15 half-hour intervals for all three quantiles. Whereas the AEMO forecast tends to yield a lower loss in the middle of the day. Moreover, the greatest difference between Models 1 and 2 is also in the middle of the day, where Model 1 tends to have a lower loss than Model 2 at 0.1 and 0.9 quantiles.

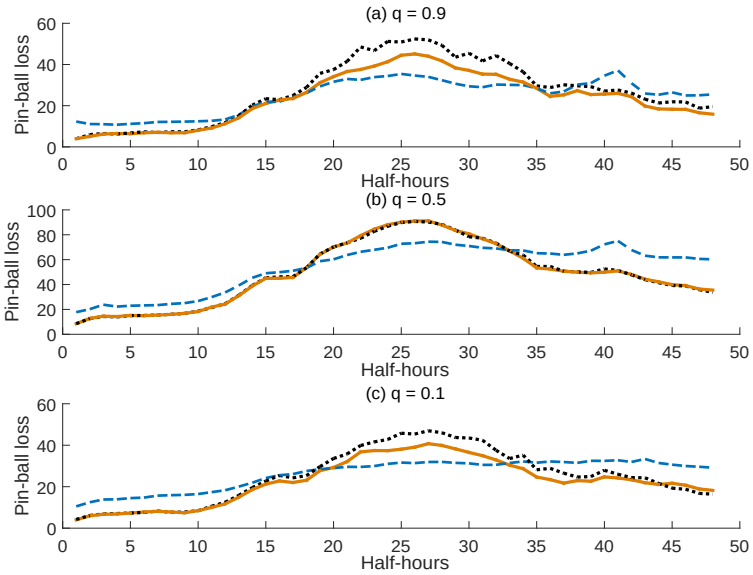

Fig. 4.2 Half-hourly pin-ball losses of the three forecasts, Model 1 (solid line), Model 2 (dotted line) and AEMO forecasts (dashed line) at quantile 0.1, 0.5 and 0.9 in Panel (a), (b) and (c), respectively. The evaluation period starts from 00:30 2014-11-24 to 00:00 2015-08-31.

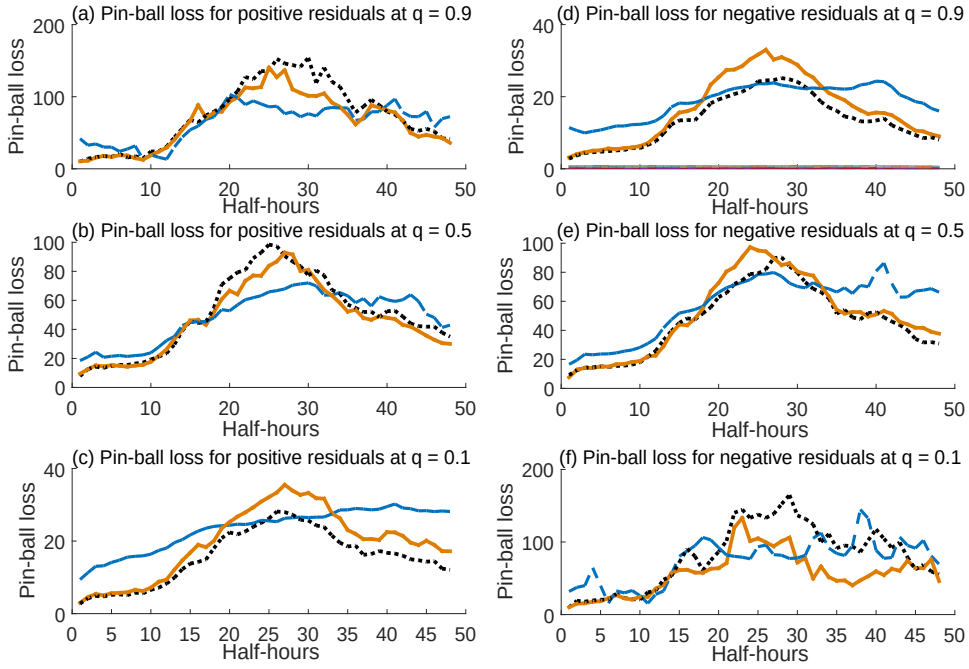

Fig. 4.3 Half-hourly pin-ball losses of the three models compared, Model 1 (solid line), Model 2 (dotted line) and AEMO forecasts (dashed line) at quantile 0.1, 0.5 and 0.9. The left (right) column are pin-ball losses for positive (negative) residuals. The evaluation period starts from 00:30 2014-11-24 to 00:00 2015-08-31.

Since the only difference between Models 1 and 2 is the type of prior residual distribution specified, to examine the connection between the types of residual distribution assumed and the differing losses shown in Figure 4.2, a further breakdown of the pin-ball losses into that for positive and negative residuals is shown in Figure 4.3. It can be seen that at the 0.9 quantile in Panels (a) and (d), the losses for Model 1 are different from the ones for Model 2, and are lower (higher) in almost all half-hour intervals when only the positive (negative) residuals are considered. The converse can also be found at 0.1 quantile. Moreover, at 0.5 quantile level, the difference is not apparent in almost all half-hour intervals. This phenomena reflects clearly the effect of the different distributional assumptions in the two models.

To visualise the effect of the different distributional assumptions imposed by the two models, consider the residual plots shown in Figure 4.4. The histograms are the plots of the residuals from the first three-year moving window estimations for the 29th half-hourly interval of a day based on a randomly chosen MCMC samples. The superimposed solid lines are the density plots of the ALD and non-parametric uniform kernel mixture density obtained from Models 1 and 2 in the first and second column, respectively. It can be seen, in Panel (a), the density plot shows that ALD assigns very small probabilities to large positive residuals when the quantile is being set at 0.9. On the other hand, considerably higher probabilities are assigned to large negative residuals. A direct consequence of this probability assignment is that large positive residual are penalised more severely than the negative ones at the same level, and the residual distribution is forced to skew toward the negative side. In Panel (b), when a flexible non-parametric residual density is used, the balance between the penalty on the level of positive and negative residuals are restored with the residual histogram becoming more symmetric in the tails. That is, the negative residuals become less extreme but at the cost of having larger positive residuals.

Conversely, at the 0.1 quantile, the non-parametric density shown in Panel (f) leads to larger negative residuals but smaller positive residuals. Based on this observation, the differences between the Pin-ball loss of Models 1 and 2 shown in Figure 4.3 becomes clear. That is, for example, when the quantile of interest is 0.9, although comparing between the densities shows in Panel (a) and (b), using the non-parametric density allows the tail and

shape behaviour on the negative side to be captured more satisfactorily. And it results in a reduction in the level of negative residuals with the corresponding pin-ball loss shown in Panel (b) of Figure (4.3) to be lower than that of Model 1. The increase in the level of positive residuals on the other hand, also leads to the pin-ball loss to be higher in Panel (a) of Figure (4.3). Since the positive residual is weighted by 0.9 under the pin-ball loss function instead of 0.1 for the negative ones, it is likely to result in a overall increase of the pin-ball loss as shown in Panel (a) of Figure 4.2.

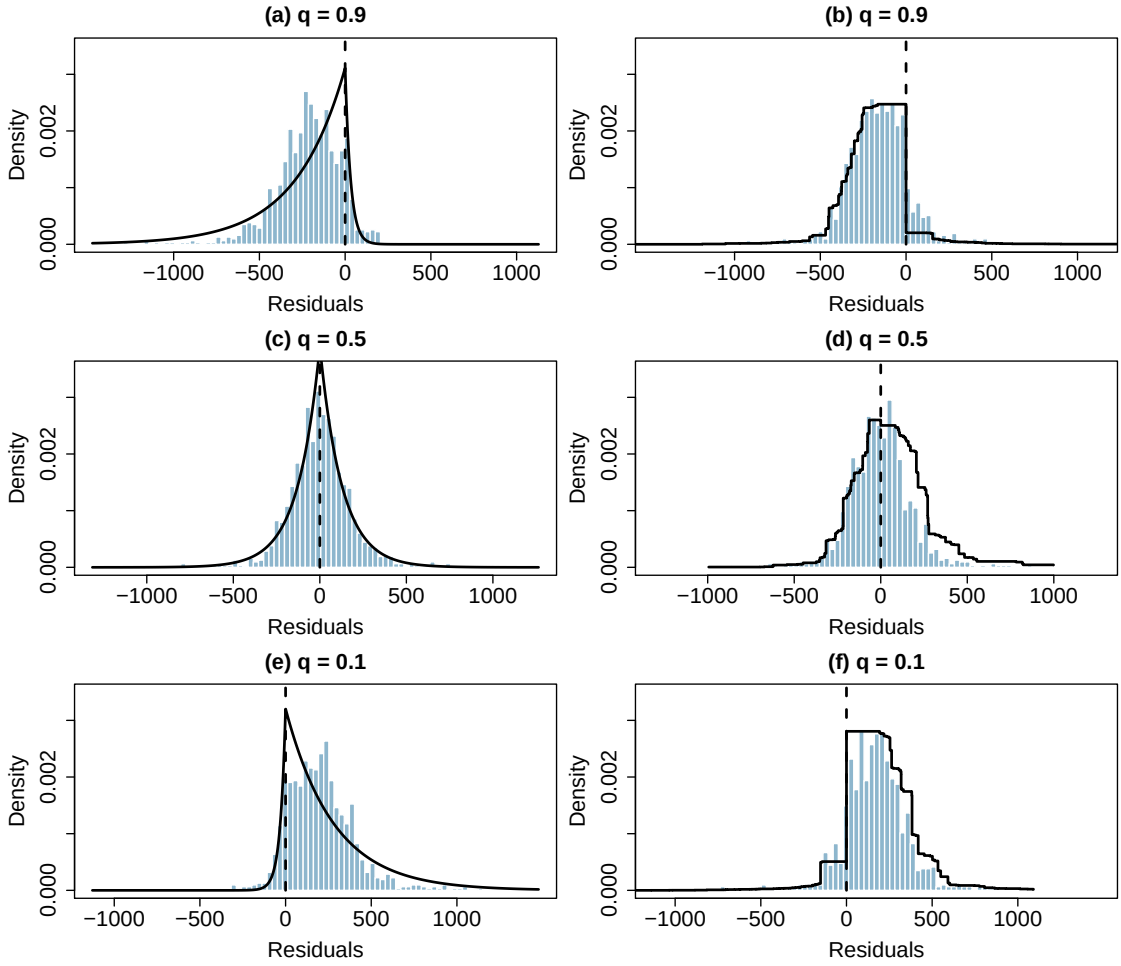

Fig. 4.4 Histogram plots of residuals from randomly selected MCMC sample based on the first three-year moving window dataset at the 29th half-hourly interval of a day with quantiles at 0.1 (bottom row), 0.5 (middle row) and 0.9 (top row), respectively. The superimposed solid lines are fitted ALD density and non-parametric density obtained from Models 1 and 2, in left and right columns, respectively.

At this point, it is important to ask which one of the models is the preferred one in general? On the surface, the results shown in Table 4.1 and Figure 4.1 indicate that Model 1 should be the choice. However, as shown in Figures 4.3 and 4.4, it is also important to consider the specific nature of the problem at hand. For example, in the context of load forecasting, Model 1 should remain as the preferred model, since if our interest is in forecasting the 0.9 quantile of load or higher quantiles, an under estimate could potentially lead to inadequate preparation for future demands and thus a possible blackout, which could be considered as more serious than the opportunity cost of having more generators ready for dispatch due to over-estimate. Consequently, the characteristics of ALD with the quantile being set at 0.9, that assign much lower probability at high level of positive residuals should be preferred even it can also lead to higher level of negative residuals comparing with using the non-parametric density.

4.6 Conclusion

Load forecasting is a problem of central interest due to its importance in the decision making process of all market participants in a deregulated electricity market. However, effective decision making in the operation of market participants requires not only an accurate point forecast but interval forecasts that could reflect variation of future load and the decision risk involved. In this chapter, two Bayesian quantile regression models are proposed for day-ahead quantile forecasts of electricity load. In the first model, ALD is used as the residual distribution with a mixture representation of asymmetric Laplace distributed random variable and a computationally efficient Gibbs sampler. The advantage of this approach is not only computational but is also robust for the reason that all parameters in the conditional distributions have closed form solutions. As a consequence, satisfactory results can be obtained even with a large number of parameters. This aspect of the method is particularly useful in terms of load forecasting due to its ability to account for complex seasonal patterns and a large number of important covariates. In the second approach, a non-parametric density is used to replace the corresponding density of ALD thus relaxing the stringent distributional

assumption of ALD. However, the results show that the performance is not preferable at least in terms of the two criteria (coverage ratio and pin-ball loss) considered.

Overall, the proposed model using ALD is both computationally efficient and very competitive in terms of forecasting accuracy when compared with the second approach and the AEMO forecast. However, the conclusion drawn is that the real preference of one model over another should be based on the nature of the problem at hand. The approach with using ALD penalises large residuals more severely in one tail than the other when the quantile of interest is different from 0.5. Whereas the non-parametric approach does not have this characteristic. Therefore, a decision on the choice of the models should be made based on whether over-prediction is preferable to under-prediction. In the context of load forecasting, the proposed model using ALD may be the preferred one.

In addition, the pin-ball loss is used as a main criterion for performance evaluation in this chapter. This criterion may not accurately reflect an individual's real attitude towards risk. Moreover, the loss function can differ based on both individuals and the nature of problems. In this chapter, it is only shown that the ALD based quantile regression is preferable in terms of pin-ball loss. The real question is however, how to flexibly incorporate different loss functions into a more general quantile regression framework, and thus achieve superior performance in terms of one specific loss function which best represents the altitude towards risk for a specific individual. This is an interesting topic which deserves future research.

Moving beyond, the focus of the thesis thus far has been on load forecasting, which is a demand side problem. This demand side problem is addressed in isolation with respect to the supply side and the general equilibrium of the market. The reason for the possibility of this isolation is due to a common characteristic of deregulated electricity markets world-wide, that is the separation of wholesale and retail prices of electricity. In the wholesale market, the price of electricity is highly variable and is set approximately every five minutes. Whereas the price in the retail market is held fixed comparatively in the short-term. Consequently, the demand, which is ultimately determined by individual consumers, is only influenced in the short-term by a comparatively fixed retail price, and is inelastic in response to the market condition changes in the wholesale market.

However, as stated previously, the success of market deregulation and the normal operation of participants in a deregulated market requires a complete knowledge of all the aspects of the market. Therefore, in the rest of this thesis, attention shifts beyond the demand side to looking at the deregulated market as a whole, in which some important issues regarding the interaction between wholesale price as the result of general equilibrium, the demand side, the behaviour of market participants on the supply side and the unique physical characteristics of the electricity are examined.

Chapter 5

The effect of transmission constraints on electricity prices

5.1 Introduction

In the National Electricity Market (NEM) of Australia, generators and consumers of electricity are spread out over a vast area and serviced by an integrated transmission network. The transmission infrastructure connects smaller areas in the market known as nodes and transfers electricity between them, in order to balance the real-time demand and supply of electricity over the entire network. As the market operator of the NEM, the Australian Energy Market Operator (AEMO) has the important role of the central coordinator, who maintains the market equilibrium by matching the bids from generators and buyers located at different locations at minimum cost. In order to reflect the local cost of electricity generation and encourage competition at all locations in the market, prices for different areas are allowed to vary, a phenomenon known as locational pricing (Ding and Fuller, 2005). In the NEM, the locational prices are set every five minutes allowing the trading activities of market participants and changes in market conditions to have almost instant effect on the prices. This variability of prices gives rise to a risk which must be managed by all market participants and consequently strategic decisions on hedging and other trading activities rely heavily on price forecasts.

Although forecasting short-term electricity prices has attracted a significant attention in the literature (see Weron, 2014, for a comprehensive survey), the price being modelled is either restricted to one locational price with explanatory variables that only represent the market conditions in that area, or an overall price level that is recorded prior to the change to locational pricing (Huisman and Mahieu, 2003; Cartea and Figueroa, 2005; Mount et al., 2006; Kanamura and Ohashi, 2008; Mount and Ju, 2014). This is in sharp contrast to the institutional setting, in which locational pricing (Ding and Fuller, 2005) that sets differential prices for smaller areas based on the local equilibrium conditions, are implemented in most deregulated markets.

Consequently, much of the existing literature on modeling prices ignores the important role of the transmission infrastructure in enabling equilibrium to be maintained at all locations on the grid. At the very least, a model of electricity prices must explicitly take into account transmission constraints which constitutes important information about the local state of the market for which forecasts are generated (Douglas and Popova, 2011; Burnett and Zhao, 2015).

This chapter explores the effect of transmission constraints on short-term electricity price variation in the Queensland region of the NEM, using detailed data about the operation of the NEM made available by the AEMO.[1] These data include five-minute dispatch reports containing the regional dispatch prices, dispatched capacities, loads, interregional flows and nodal constrains together with bid reports that provide information on the bidding process for the previous day. The use of five-minute data on dispatch prices, instead of the half-hourly trading prices commonly studied in the literature, has the advantage that they are the actual prices produced by the dispatch algorithm and therefore preserve the potential causes of electricity price variations without contamination by averaging. In this way a clearer identification of the cause of variations in trading prices is facilitated. Another innovative aspect of the dataset is the inclusion of comprehensive bid information, which contains the bids from all generators in the market at the same five-minute frequency. The bid information enables the construction of a variable, the changes in bid quantities, which represents the

[1] See www.nemweb.com.au

changes of the supply side market condition and is used in the regression analysis in order to facilitate the identification of the transmission constraints effect. The main contribution of this chapter is to use quantile regressions to identify the effect of transmission constraints on slices of the price distribution. The result indicates that transmission constraints are an important cause of short-term price variation, a result that emphasises its importance to any model that seeks to forecast regional electricity prices in an integrated electricity market.

5.2 Institutional background

The NEM, introduced in December 1998, operates one of the world's largest interconnected electricity systems which comprises five regions, namely New South Wales, Victoria, Queensland, South Australia and Tasmania. The different regions of the NEM are connected by inter-regional transmission lines. Each region is divided into a number of nodes which are connected via the nodal transmission network. Wholesale trading in the NEM is conducted as a spot market where supply and demand in each node is instantaneously matched through a centrally-coordinated dispatch process managed by the AEMO. Generators are allowed to bid generation capacity within 10 price bands which fall between the floor price of -1,000 AUD/MWh and the cap price of 13,100 AUD/MWh. The central dispatch algorithm uses the current bid information to schedule generation in order to satisfy the demand for electricity at minimum cost, taking into account the security of operation of the entire system subject to certain physical constraints.

The equilibrium spot price of electricity is determined as the outcome of the dispatch process. The preliminary nodal price is calculated for each node in every five-minute interval, based on the real-time nodal equilibrium conditions. The five-minute regional dispatch price, which, subject to minor adjustments, is given by the highest nodal price recorded in the region. For settlement purpose, a half-hourly regional trading price is quoted, which is the average of the six regional dispatch prices in the half-hour trading interval. For each region, the regional trading price and dispatch price are reported by the AEMO.

A time series plot of dispatch prices for Queensland in the period from 30 November 2012 to 7 November 2013 ($T = 98784$ observations) is shown in Panel (a) of Figure 5.1. The most striking feature of this time series is the extreme volatility. The average dispatch price for the period is 69 AUD/MWh, but during the period under consideration the dispatch price exceeded 1,000 AUD/MWh 206 times, 12,000 AUD/MWh 27 times and fell below -900 AUD/MWh 11 times. Furthermore, in many instances, the abnormally high price events occurred immediately after a dispatch interval in which price was broadly speaking at a normal level and returned to a normal level in the dispatch interval immediately thereafter. Panel (b) of Figure 5.1 is a scatter plot of dispatch prices and loads for the same sample period. It shows that abnormal price events occur over a wide range of load and not merely at very high levels where the system could reasonably be characterised as being at capacity.

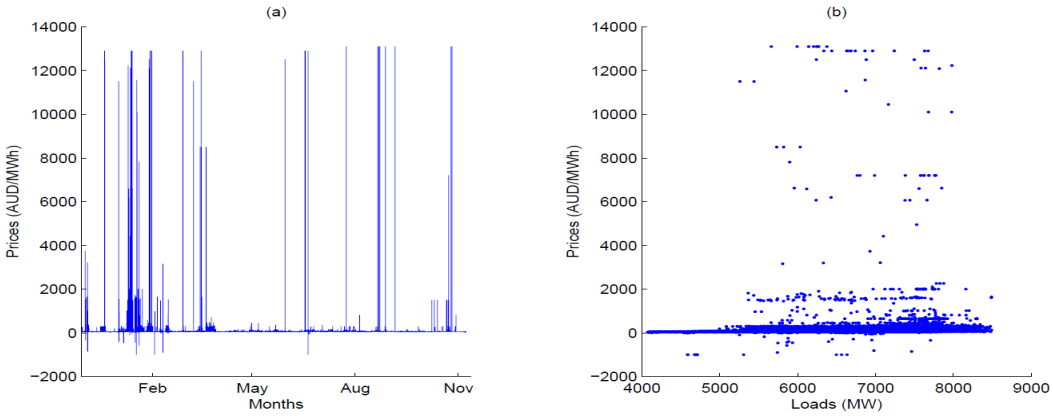

Fig. 5.1 Panel (a) shows a time series plot of dispatch prices while Panel (b) shows a scatter plot of dispatch prices and loads. The data are for five-minute intervals in the Queensland region during the period 30 November 2012 to 7 November 2013, $T = 98784$ observations.

The empirical evidence provided by Figure 5.1 suggests that lagged price and load alone cannot adequately explain the behaviour of dispatch price. These covariates are not, therefore, able to account for the extreme variation in price which is a characteristic of this sample

period and of dispatch price more generally and it is necessary to explore more formally the role of transmission constraints in explaining the volatility of dispatch price.

In scheduling the dispatch strategy, the dispatch algorithm is designed to fulfil electricity demand for the whole market at minimum cost. The cost is measured based on the bid prices of dispatched marginal generators. If dispatch prices were different on the two ends of a transmission line, the higher price at one end is the consequence of dispatching generator in the area with higher bid price. Given there is extra transmission capacity available on the line, the overall cost can be reduced by dispatching generation resource at the end with a lower bid price and transferring the electricity into the end with higher price.[2] Therefore, without transmission constraints, regional prices between neighbouring regions should be relatively close. On the other hand, the occurrence of transmission constraints in the network inhibits the flow of electricity across the market, and the result can be differing market conditions and potentially large differences between regional prices.

In general, transmission constraints can be classified into two types. *Regional constraints* are defined as constraints that occur on the interconnectors across regional boundaries. The occurrences of regional constraints can eliminate the generators located in other regions that may otherwise be available to the dispatch algorithm. In the event of a demand increase in a constrained region, the dispatch algorithm can only rely on dispatching generators in the constrained region with possibly higher bid price. Thus, regional constraints can lead to a higher or more volatile regional price for the constrained region. *Nodal constraints* are defined as constraints that occur on the transmission lines between nodes in the same region. Nodal constraints can potentially isolate some nodes within a region. In these isolated nodes, consumption has to be fulfilled by only relying on local generation resources regardless of the possibly higher bid price compared with that of available generators at the other side of the constraints. Thus, the nodal prices at constrained nodes can become higher or more volatile, which in turn can potentially set the price for the whole region due to the uniform regional pricing rule. Accordingly, the differences between regional prices can become large when either regional or nodal constraints (or both) occur.

[2]Transmission loss, which is only a small proportion of the total price, is ignored.

Since the Queensland region is the focus of this chapter, it is useful to spell out the Queensland network in a little detail. Queensland is connected to the rest of the NEM by two transmission lines, QNI (Queensland New South Wales Interconnector) and Directlink, that cross the border between Queensland and New South Wales. The import capacity of QNI and Directlink into Queensland is around 150 MW and 20 MW respectively. The exact capacity varies with many factors that are related to the physical characteristics of the network and the market in general. Within Queensland itself, the main transmission lines are located along a long strip of the coastal area where the majority of the population is located. This configuration means that any occurrence of nodal constraints in the region are likely to separate between the north and the south of the region and completely isolated the north from the rest of the market.

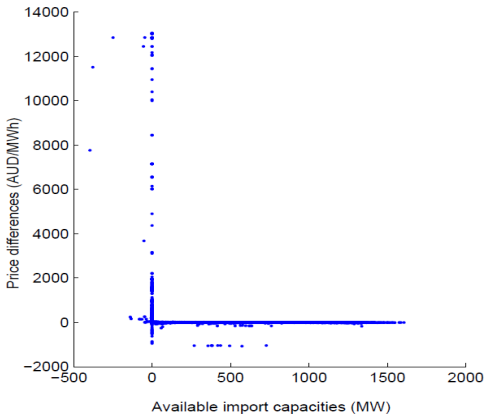

Fig. 5.2 Scatter plot of the regional price differences between Queensland and New South Wales (Queensland price minus New South Wales price) and available import capacities on the inter-regional transmission lines for Queensland. The data are for five-minute intervals during the period 30 November 2012 to 7 November 2013, $T = 98784$ observations.

Casual empirical evidence for the effect of transmission constraints on setting the regional price of Queensland can be obtained by plotting the price differences between Queensland and New South Wales (Queensland price minus New South Wales price) against available import capacities for Queensland. For the import capacities, the sum of available import

capacities on the two interconnectors into Queensland is used. The available import capacity for each line is the import limit minus the amount of actual electricity flow and, in the case of a constraint on export, the available import capacity is the sum of import and export limits on the line. An available import capacity smaller than 0.1 MW is treated as a transmission constraint. The resulting plot is shown in Figure 5.2. The effect of constraints is clearly displayed because the price difference only becomes positive and large when the available import capacity for Queensland is constrained. In fact, Figure 5.2 demonstrates that the existence of a regional transmission constraint is a necessary condition for abnormally large price differences between Queensland and New South Wales.

A more detailed illustration of the effect of transmission constraints can be shown by classifying the price differences according to their correspondence with nodal or regional constraints in Queensland. To set a benchmark, all the observations of the difference in prices are plotted against inter-regional flows in Panel (a) of Figure 5.3. It shows that large price differences mostly occur at the inflow level of approximately 220 MW, which implies the import capacity for Queensland is around this level. Comparing this import capacity with the usual system load of around 6500 MW for Queensland, it shows the limited ability of alleviating the unexpected load changes in Queensland by means of the two interconnectors.

To highlight the effect of regional constraints, price differences without regional constraints are plotted in Panel (b) of Figure 5.3. The maximum price difference without regional constraints is around 30 AUD/MWh. Compared with the maximum difference of around 13,000 AUD/MWh in Panel (a), the strong correlation between extreme price differences and regional constraints in Queensland is again apparent. To highlight the effect of nodal constraints, the price differences in the absence of nodal constraints are plotted in Panel (c) of Figure 5.3. It can be seen that most extreme price differences are eliminated; an indication of the importance of nodal constraints in explaining regional prices.

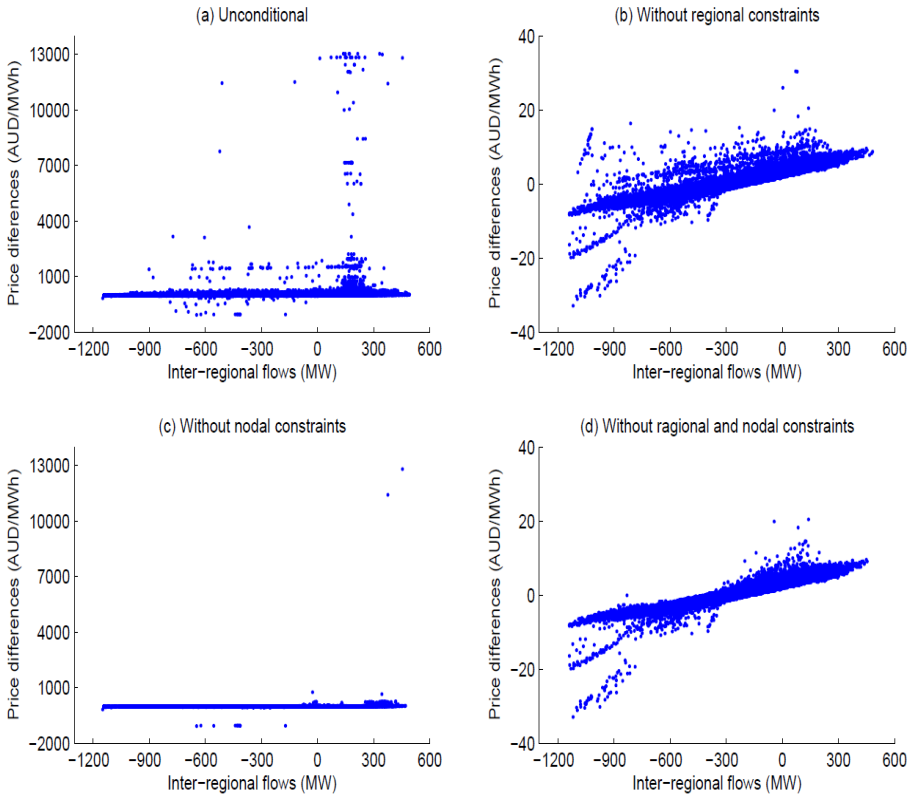

Fig. 5.3 The price differences between Queensland and New South Wales plotted against inter-regional flows from New South Wales into Queensland. All the price differences, price differences without regional constraints, price differences without nodal constraints and price differences without any constraints are shown in Panels (a), (b), (c) and (d), respectively.

In addition, there are instances, shown most clearly in Panel (b) of Figure 5.3, of positive price differences at negative inter-regional flows (electricity is flowing to New South Wales, but Queensland has a higher price). These instances are due to nodal constraints in Queensland only. To see if the positive price differences at negative flows are indeed related to nodal constraints, the observation of price differences without both regional and nodal constraints are plotted in Panel (d) of Figure 5.3. It is clear that all the positive price differences at negative flows due to nodal constraints are eliminated. Moreover, without both constraints, all the instances of extreme price difference disappear. The positive relationship between price differences and inter-regional flows shown in Panel (d) indicates the magnitude of

contribution of transmission loss to the regional price differences, which only ranges from -40 to 20 AUD/MWh.

5.3 Preliminary analysis

The informal evidence summarised in Figure 5.3 is strongly suggestive that both nodal and regional constraints are pivotal determinants of regional electricity prices. The task is now to explore this hypothesis in a more formal setting using regression analysis.

5.3.1 Explanatory variables

The dependent variable in a preliminary regression analysis of the effect of constraints on electricity prices is taken to be the five-minute Queensland regional dispatch price, P_t. Although the time series of an electricity price is known to be a persistent series, it is also generally accepted that it is stationary (Christensen et al., 2009; Weron, 2014) so that the lagged value of price, P_{t-1}, is likely to be a significant explanatory variable.

It is also likely that both demand and supply side influences will play a role in determining prices. On the demand side, the simplifying assumption is made that the effect of the level of load is accurately captured by the persistence in prices. Thus, the change in the dispatched quantity, ΔD_t becomes an important explanatory variable. In fact, the change (instead of the level) of demand is preferred, since the information shown in Panel (b) of Figure 5.1 indicates that the level of demand does not play an important role in representing current market conditions. Similarly, to capture the supply side changes in the market condition, another variable representing the changes in bid quantities, ΔB_t is constructed. In doing so, the bid prices and quantities from every generator are obtained from the 'Yesterdays Bids Reports' at a five-minute frequency. Then these bids are grouped into different regions according to their generator ID.[3] Adding the bids from the generators in Queensland in merit order, the five-minute bid curves for the region are obtained. By substituting the current

[3] See, http://www.aemo.com.au/Electricity/Market-Operations/Loss-Factors-and-Regional-Boundaries

dispatch price obtained from the dispatch report into the current bid curve, the quantity of bid available at the current dispatch price is produced. Then substituting the current price into the bid curve of the previous dispatch interval, a bid quantity available at the current price level in the previous bid condition is obtained. The change between these two bid quantities represents the change of supply condition in the most relevant price level and is the change in bid quantities finally obtained. A times series plot of the two constructed variables is shown in Figure 5.4. In general, it is clear that the changes in dispatched quantities and bids are quite random and moderate compared with the extreme variation of the dispatch price, shown in Figure 5.1.

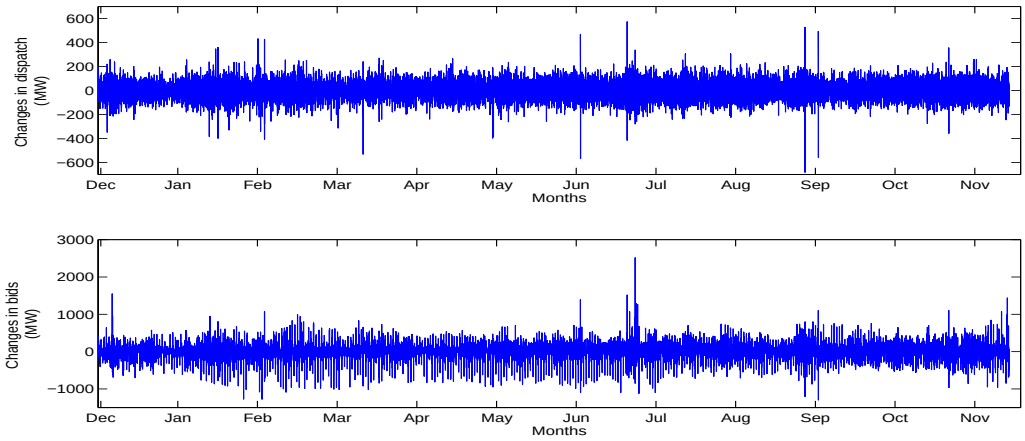

Fig. 5.4 Time series plots of changes in dispatched quantities and changes in bid quantities for Queensland from 30 November 2012 to 7 November 2013, $T = 98784$ observations.

Of course the primary interest in this regression analysis is the effect of transmission constraints on prices. Accordingly a dummy variable, C_t, is constructed that takes the value 1 if any constraint is present and 0 otherwise. The information required to construct this dummy variable is contained in the dispatch reports. Nodal constraints are reported directly but regional constraints need to be inferred based on the import capacities and actual flows on the interconnectors between Queensland and New South Wales.

In Figure 5.5, a time series plot of the dummy variable for constraints is shown together with the times series plot of dispatch prices. The much more frequent occurrence of transmis-

sion constraints compared with that of extreme prices indicates that transmission constraints may not be the sufficient condition for their occurrence. In other words, a transmission constraints dummy variable for the whole region is not adequate for modelling the detailed behaviour of prices. This task will require detailed knowledge of constraints, demand changes and the current dispatch and bid profiles at each specific node.

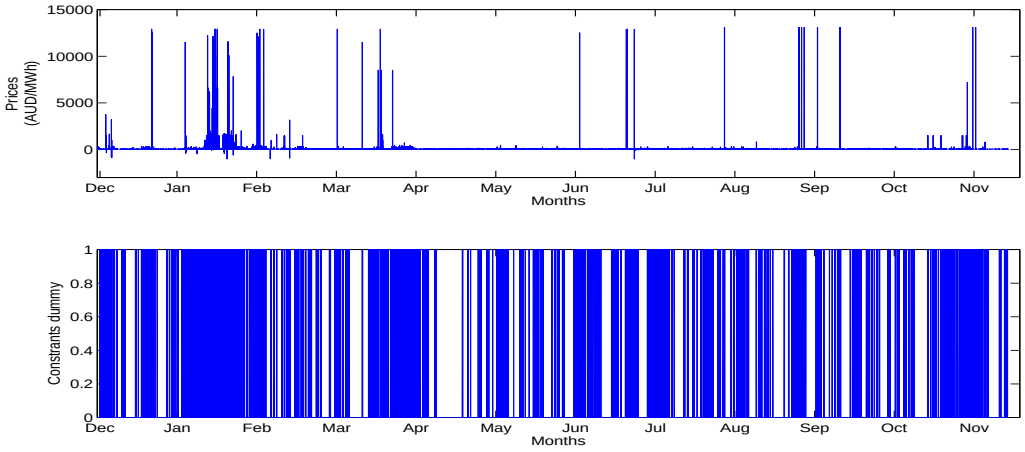

Fig. 5.5 Time series plots of the regional dispatch price and constraints dummy variable for Queensland from 30 November 2012 to 7 November 2013, $T = 98784$ observations.

Table 5.1 Summary statistics for dispatch prices with or without transmission constraints, in Queensland from 30 November 2012 to 7 November 2013.

	Prices with $C_t = 0$	Prices with $C_t = 1$
50%	52.99	64.94
75%	57.00	99.85
90%	60.59	217.39
95%	66.09	299.99
99%	93.63	1565.9
Maximum	439.61	13100
Mean	54.21	156.48
Standard deviation	15.56	686.71
Observations	84027	14756

Table 5.1 reports summary statistics of the prices conditioned on the constraints dummy variable. The sample quantiles of price suggest that while the median only increases slightly, transmission constraints can have a large impact on the higher quantiles of price. That is, the existence of a transmission constraint may not be a sufficient condition for the occurrence of an extreme price event, but it does appear to be a necessary condition. Without constraints, the maximum price is only 500 AUD/MWh. Moreover, the much higher standard deviation in price with constraints also shows that the price becomes more volatile when the network is constrained.

Another important aspect of the prices shown in Table 5.1 is that the maximum recorded price level with $C_t = 1$ is the current maximum allowable price of 13,100 AUD/MWh. The presence of these extreme price observations is likely to have a noticeable impact on the mean as a measure of the central tendency of the price distribution. In situations such as these, sizeable differences between the mean and median are to be expected, preliminary evidence of which can be seen by comparing the median price with constraints (64.94 AUD/MWh) to the mean price with constraints (156.48 AUD/MWh).

5.3.2 Regression models

Given the variables introduced in Section 5.3.1, a simple but effective way to identify the effect of transmission constraints on price is to specify the price, P_t as having the linear form:

$$\begin{aligned} P_t =& \beta_1 + \beta_2 P_{t-1} + \beta_3 \Delta D_t + \beta_4 \Delta B_t \\ &+ \beta_5 C_t + \beta_6 C_t \times P_{t-1} + \beta_7 C_t \times \Delta D_t + \beta_8 C_t \times \Delta B_t + u_t, \end{aligned} \tag{5.1}$$

where, u_t is the residual term. In this specification, the price is determined by three fundamental covariates: lagged price, P_{t-1}; the change in dispatched quantity, ΔD_t, which captures changes on the demand side of the market; and the change in bids quantity, ΔB_t, which represents supply side market changes. The expected signs for the relevant coefficients are $\beta_2, \beta_3 > 0$ and $\beta_4 < 0$. In addition, the dummy variable for constraints, C_t and its interaction with the three covariates allow the specification to capture the effects of constraints on the

behaviour of price. If transmission constraints play a significant role in determining the price, then at least one of β_5, β_6, β_7 and β_8 will be statistically significant. It is expected that $\beta_5, \beta_7 > 0$ and $\beta_6, \beta_8 < 0$ meaning that the price is generally higher or more volatile with constraints.

The specification in (5.1) is estimated in three ways.

(i) In order to provide a benchmark, ordinary least squares is used, in which the estimated coefficients represent the mean effect of covariates over the entire price distribution.

(ii) The impact of the extremely high price, indicated by the significant difference between the mean and median prices shown in Table 5.1, makes it necessary to implement a procedure that is robust to outliers. Consequently, Equation (5.1) is also estimated by means of an M regression with the bi-square objective function:

$$\rho(z_t) = \begin{cases} \frac{c^2}{6}\left(1 - \left(1 - (\frac{z_t}{c})^2\right)^3\right), & \text{if} \quad |z_t| \leq c, \\ \frac{c^2}{6}, & \text{if} \quad |z_t| > c, \end{cases}$$
$$z_t = \frac{u_t}{s},$$

where u_t is the residual term, s is an iteratively estimated scale parameter and c is a tuning constant. The M regression is equivalent to the maximum likelihood approach with the bi-square objective function that puts 0 weight on extreme outliers of residuals and it is therefore more robust to outliers than using the normal density function or its equivalent ordinary least squares procedure.

(iii) An alternative robust estimation procedure which also has the advantage of focusing attention on the median effect of constraints is a median regression. The objective function for the median regression is given by

$$\rho_\tau(u_t) = u_t(\tau - I(u_t < 0)). \tag{5.2}$$

By minimising the sum of the objective function, $\rho_\tau(u_t)$ in Equation (5.2) with $\tau = 0.5$ and evaluating at u_t, the residuals obtained using the specification in Equation (5.1), the estimates are mainly determined by the order of the residuals instead of by their values in the tails of the distribution. In this way median regression ensures that the parameter estimates are robust to outliers.

Table 5.2 Ordinary least squares, M and median regressions of the model in Equation (5.1). The dependent variable is the five-minute dispatch price in Queensland from 30 November 2012 to 7 November 2013, $T = 98784$ observations. The values in parentheses are t statistics.

	OLS	M	Median
Intercept	53.58	0.75	4.71
	(131.16)	(24.97)	(2.90)
P_{t-1}	0.01	0.98	0.91
	(2.20)	(1715.99)	(29.57)
ΔD_t	0.02	0.01	0.01
	(6.69)	(60.62)	(38.32)
ΔB_t	−0.01	−0.01	−0.01
	(−9.05)	(−57.24)	(−40.26)
C_t	64.88	108.49	37.96
	(6.69)	(1140.45)	(9.98)
$C_t \times P_{t-1}$	0.23	−0.71	−0.57
	(4.08)	(−1206.40)	(−10.27)
$C_t \times \Delta D_t$	0.41	0.44	0.04
	(1.40)	(288.87)	(4.44)
$C_t \times \Delta B_t$	−0.75	−0.69	−0.04
	(−4.36)	(−565.94)	(−21.42)

The results of estimating the model in (5.1) by these three methods are reported in Table 5.2. In general, the signs of all the coefficients for the behaviour of price in the absence of constraints are in line with expectations. The signs on the demand and supply variables

accord with intuition ($\beta_3 > 0$ and $\beta_4 < 0$) and are relatively stable across the three estimation methods. But the same cannot be said for β_2, the coefficient on P_{t-1}. The recorded estimates range from 0.01 to 0.98 with the robust procedures, M and median regression, yielding estimated coefficients for P_{t-1} that are much higher than the ordinary least square estimate. This is due the fact that if the serial correlation is high, the resulting extreme negative residuals after price spikes are penalised more severely in the ordinary least squares procedure than in the robust alternatives.

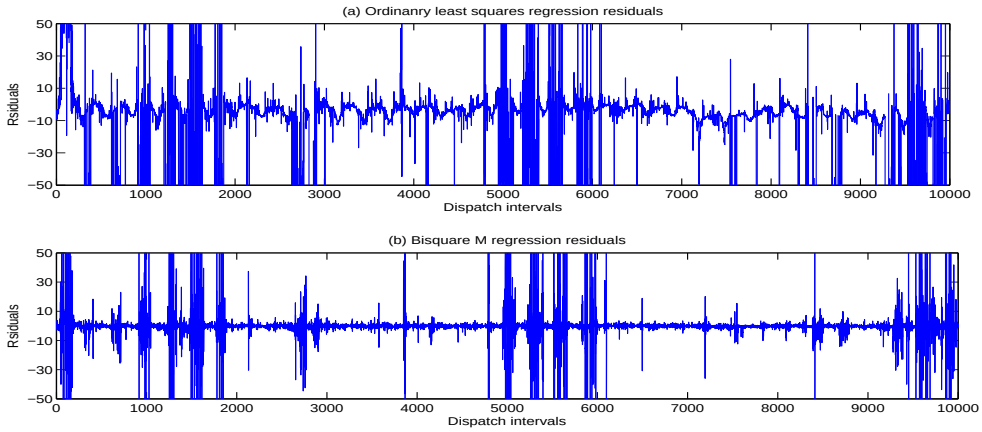

Fig. 5.6 Time series plots of residuals obtained from ordinary least squares and bisquare M regressions, for the first 10,000 dispatch intervals. The level of residuals being displayed is restricted to be between -50 and 50 for a clearer presentation.

To investigate this issue further, consider the time series plot of residuals for a representative period obtained from both ordinary least squares and M regression shown in Figure 5.6, in which it is quite clear that the residuals plotted in top panel (ordinary least squares) have significant remaining residual autocorrelation while those in the lower panel (M regression) do not. It is the absence of a strong autoregressive parameter when the model is estimated by ordinary least squares that induces the periodic pattern in the residuals. Furthermore, the results in Table 5.2 for the M and median regressions indicate that price is indeed persistent in the central mass of the price distribution when outliers are treated correctly; a result that is consistent with commonly acknowledged high level of persistence in electricity prices (Knittel and Roberts, 2005; Weron and Misiorek, 2008).

Turning now to the effect of constraints, there is one general conclusion to be drawn. The coefficients involving the constraints, β_5, β_6, β_7 and β_8 in Equation (5.1) are statistically significant across all the methods. The sole exception is the estimate of β_7, the interaction of the constraint dummy with the demand side variable, when estimated by ordinary least squares. However, an F-test based on the ordinary least squares procedure with the hypothesis $\beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$ yields the result $F_{(4,98773)} = 707.13$ with a p value of 0.00. These result provides overwhelming evidence that the price of electricity exhibits fundamentally different behaviour when constraints are present.

Moving to the specifics of the estimation results when constraints are present, the first thing to note is that the estimates of the coefficient β_5 are positive for all three procedures. This means that the expected or median price of electricity is significantly higher when constraints are present. It is no surprise, given the pattern of values recorded for β_2 and the residual plots in Figure 5.6, that the estimates of β_6, the parameter on the interaction term $C_t \times P_{t-1}$ show significant differences between the robust and non-robust estimation procedures. The expectation is that this coefficient has a negative value indicating a drop in persistence in the presence of constraints and this is exactly what results when robust estimation methods are used.

5.4 Modelling the quantiles

The M regression and median regression results reported in columns 3 and 4 of Table 5.2, respectively, point to a fact that the mean and median effects of constraints implied by the estimated coefficients for C_t, $C_t \times \Delta D_t$ and $C_t \times \Delta B_t$ are different. In particular, the coefficients returned by median regression imply that the impact of constraints on the median price are smaller than the corresponding impact on the mean price given by the M regression. This difference in estimated effect of transmission constraints indicates that a fertile area of inquiry would be an examination of the effects of constraints at other quantiles of the price distribution.

To explore the effects of transmission constraints at other quantiles, a series of quantile regressions based on the quantiles of the price, $\tau \in \{0.5, 0.75, 0.9, 0.95, 0.99\}$, are implemented using the same variables listed in Table 5.2. The models are specified to have a linear form:

$$\begin{aligned} P_t =& \beta_{1\,\tau} + \beta_{2\,\tau} P_{t-1} + \beta_{3\,\tau} \Delta D_t + \beta_{4\,\tau} \Delta B_t \\ &+ \beta_{5\,\tau} C_t + \beta_{6\,\tau} C_t \times P_{t-1} + \beta_{7\,\tau} C_t \times \Delta D_t + \beta_{8\,\tau} C_t \times \Delta B_t + u_t, \end{aligned} \tag{5.3}$$

Table 5.3 Quantile regressions and M regression for the effect of transmission constraints with quantiles, $\tau \in \{0.5, 0.75, 0.9, 0.95, 0.99\}$. The dependent variable is the five-minute dispatch price in Queensland from 30 November 2012 to 7 November 2013. The values in parentheses are t statistics based on a robust estimate of the covariance matrix.

	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$	$\tau = 0.95$	$\tau = 0.99$	M
Intercept	4.71	2.17	1.62	2.10	10.03	0.75
	(2.90)	(2.88)	(1.06)	(0.75)	(41.39)	(24.97)
P_{t-1}	0.91	0.96	0.99	1.01	1.01	0.98
	(29.57)	(69.58)	(34.57)	(19.69)	(496.34)	(1715.99)
ΔD_t	0.01	0.01	0.02	0.02	0.07	0.01
	(38.32)	(53.26)	(29.89)	(60.26)	(59.05)	(60.62)
ΔB_t	−0.01	−0.01	−0.01	−0.01	−0.02	−0.01
	(−40.26)	(−47.09)	(−58.54)	(−64.35)	(−58.69)	(−57.24)
C_t	37.96	4.26	30.97	41.36	309.02	108.49
	(9.98)	(3.46)	(4.97)	(10.20)	(27.68)	(1140.45)
$C_t \times P_{t-1}$	−0.57	0.01	0.02	0.64	3.85	−0.71
	(−10.27)	(0.82)	(0.42)	(12.56)	(612.19)	(−1206.40)
$C_t \times \Delta D_t$	0.04	0.07	0.27	0.37	0.44	0.44
	(4.44)	(11.73)	(23.99)	(11.24)	(25.86)	(288.87)
$C_t \times \Delta B_t$	−0.04	−0.07	−0.19	−0.31	−1.02	−0.69
	(−21.42)	(−11.79)	(−20.04)	(−26.72)	(−42.59)	(−565.94)

The results obtained from this series of quantile regressions together with t statistics based on a robust covariance matrix estimator of Powell (1991) are reported in Table 5.3 together with the repeated M regression results for comparative purposes.

There are a number of noteworthy observations to be made based on the results in Table 5.3. Probably the most striking one is the implied coefficient on lagged price in the presence of constraints, $1.01 + 3.85 = 4.86$. The occurrence of the constraint imparts a significant change in the skewness of the price distribution and serves to emphasise the importance of constraints in modelling price. This behaviour in the tails of the distribution is still consistent with a stationary price process provided that occasional negative price shocks occur (Koenker and Xiao, 2006).

The estimated coefficients for the constraint dummy variable are positive and significant across all the quantiles with sharp increases in the upper tail of the distribution. The effect of $C_t \times \Delta D_t$ and $C_t \times \Delta B_t$, rows 8 and 9 also differ from that without constraints, rows 4 and 5 by an order of magnitude. These results are not surprising because, in the presence of constraints, any change in the market conditions can only be balanced by local generation capacity.

One irregularity shown in Table 5.3 is that the constant effect of constraints does not increase monotonically when moving into the higher quantiles. This contradicts the intuition of the positive relationship between constraints and the price level. To investigate this in a little more detail, Figure 5.7 shows the estimated parameters, together with their 95% confidence intervals, over 50 equally spaced quantiles ranging from 0.01 to 0.99. To account for the time-series feature of the data, the percentage confidence interval based on the block bootstrap (Liu and Singh, 1992; Fitzenberger, 1998) is used. Specifically the seasonal block bootstrap (Politis, 2001; Chan et al., 2004) is adopted with block size of 288, which corresponds to the daily seasonal pattern commonly observed in electricity price data (Weron et al., 2004). The block bootstrap was also implemented with an optimal block size of around 960 obtained using the procedure of Patton et al. (2009). The results obtained from the latter method were almost identical to those computed using the seasonal block bootstrap and so are omitted. Compared with the t statistics reported in Table 5.3, the block

bootstrap leads to more conservative rejection decisions. Moreover, since the price data under consideration include extreme price events that are largely unpredictable given the data at hand, this nonparametric resampling procedure is preferred to a parametric approach as used, for example, by Mount and Ju (2014).

Figure 5.7 demonstrates that the price distribution changes not only in location and scale but also in shape in response to the changes in covariates. The intercept decreases as moving into the higher quantiles, while the coefficient on lagged price increases. The latter phenomenon has already been noted. In Panels (c) and (d), the shape changing effects of the covariates on price distribution become more pronounced after the 80% quantile, with the effect of the demand influence becoming more positive and that of the supply variable more negative.

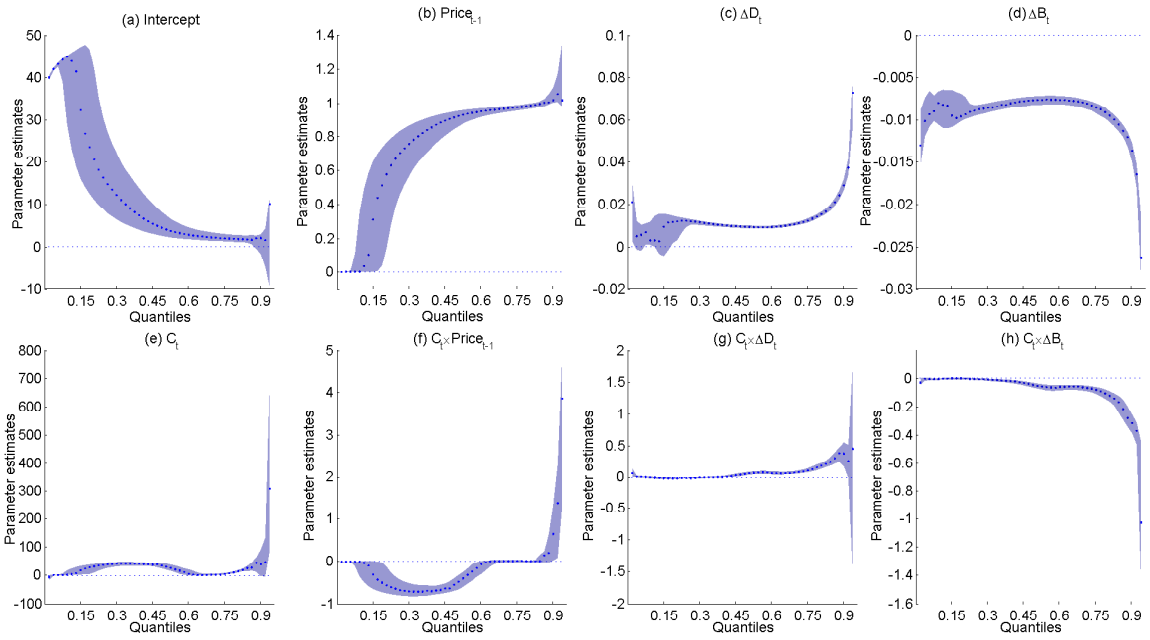

Fig. 5.7 Estimated parameter values from quantile regressions of the form given in Equation (5.3) for 50 equally spaced quantiles from 0.01 to 0.99.

In terms of the coefficients on the constraint dummies, the constant effect of constraints, C_t in Panel (e), shows that constraints contribute positively to price level in a wide range of quantiles. Note however that there is a range over which the effect is quite small. This pattern reflects the fact that when the highest price without constraints (about 400 AUD/MWh

as shown Table 5.1) is reached, there is little additional information in the data about the effect of constraints because these higher prices are always accompanied by transmission constraints. In Panel (f), the effect of the lagged price in the presence of constraints is initially negative and reaches 0 at around the 75% quantile. This pattern is almost a mirror image of the constant effect of constraints shown in Panel (e). In fact what emerges from Panels (e), (f), (g) and (h) of Figure 5.7 is that the variation and uncertainty of price in the presence of constraints becomes increasingly explained by demand and supply influences, the coefficients on $C_t \times \Delta D_t$ and $C_t \times \Delta B_t$.

The hypothesis that transmission constraints have a positive effect on price, or in other words that the coefficient on C_t in Panel (e) of Figure 5.7 is non-negative, may be tested formally in terms of a Wald-type test statistic based on a specific quantile. If a 95% confidence interval is used and each quantile is examined separately, the Wald test statistics give information that is similar to those shown in Figure 5.7 but with (less realistic) symmetric confidence interval. Rather than pursuing this strategy, tests for location shift, location-scale shift or positive effects of a covariate over a wide range of quantiles can be formulated using the Kolmogorov-Smirnov (KS) type statistic (Koenker, 2005). Let $\mathscr{T}$ be the closed interval $[0.1, 0.9]$ which defines the quantile range used for testing, then three null hypotheses of interest are:

$$
\begin{aligned}
&H_0 : \beta_{5\,\tau}, \beta_{7\,\tau}, \beta_{8\,\tau} = 0, && \tau \in \mathscr{T} \quad \text{[No effect]};\\
&H_0 : \beta_{5\,\tau} = \beta_5, \beta_{7\,\tau} = \beta_7, \beta_{8\,\tau} = \beta_8, && \tau \in \mathscr{T} \quad \text{[Constant effect]};\\
&H_0 : \beta_{5\,\tau}, \beta_{7\,\tau}, \beta_{8\,\tau} \geq 0, && \tau \in \mathscr{T} \quad \text{[Positive effect]}.
\end{aligned}
$$

To account for the time-series feature of the data, the block bootstrap version of the test proposed by Chernozhukov and Fernández-Val (2005) is used with a similar resampling scheme to that employed in constructing the confidence intervals in Figure 5.7. The test results are reported in Table 5.4.

Table 5.4 Kolmogorov-Smirnov tests of the null hypotheses of no constraint effects, constant constraint effects and positive constraint effect on price. Critical values are based on 10,000 block bootstrap replications with block size of 288.

Hypothesis	Null	KS statistics	95% critical value	Decision
No effect	$\beta_{5\tau} = 0$	11.89	3.77	Reject
	$\beta_{7\tau} = 0$	6.52	3.51	Reject
	$\beta_{8\tau} = 0$	5.56	3.15	Reject
Constant effect	$\beta_{5\tau} = \beta_5$	5.74	2.72	Reject
	$\beta_{7\tau} = \beta_7$	1.35	2.06	Do Not Reject
	$\beta_{8\tau} = \beta_8$	4.01	2.04	Reject
Positive effect	$\beta_{5\tau} \geq 0$	-0.58	3.03	Do Not Reject
	$\beta_{7\tau} \geq 0$	3.30	2.87	Reject
	$\beta_{8\tau} \geq 0$	5.56	3.11	Reject

The null hypothesis of zero quantile effects of transmission constraints on price is rejected decisively in all three tests. The constant quantile effect is rejected for C_t and $C_t \times \Delta B_t$. Finally, a positive quantile effect is rejected for $C_t \times \Delta D_t$ and $C_t \times \Delta B_t$. These results are not surprising and generally align with the patterns shown in Figure 5.7, and further confirm the significance of the effect of constraints on price.

The block bootstrap procedure used for testing the quantile effect of constraints requires the bootstrapped samples to be obtained by randomly drawing from the historical observations. This procedure is therefore similar to the commonly used procedure, known as *historical simulation*, for calculating Value-at-Risk (Jorion, 2007), where Value-at-Risk is defined as the amount a firm can lose with a given probability and a fixed time horizon. Thus when it is applied to electricity market (Chan and Gray, 2006), Value-at-Risk becomes a representation of the risk involved in the routine trading activities of electricity market participants.

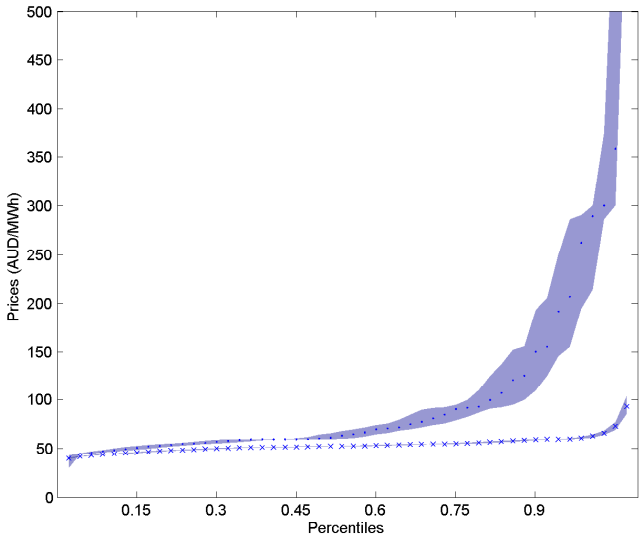

Fig. 5.8 The price levels at 50 equally spaced quantiles based on the simulated price from the bootstrap procedure. The medians of the simulated quantiles of price with and without constraints are shown as dots and crosses, respectively, with the corresponding 95% intervals indicated as shaded area.

In the context of exploring the effect of transmission constraints, it becomes a natural question to ask whether simulated prices conditioned on constraints can have a distribution that is significantly different from one which is not conditioned on constraints. Figure 5.8 shows 50 equally spaced quantiles (from 0.01 to 0.99) of the simulated price from the bootstrap procedure. The medians of the simulated quantiles of price conditioned on constraints and when constraints are absent are shown as dots and crosses, respectively, with the corresponding 95% confidence intervals indicated as shaded area. The scale on the vertical axis of Figure 5.8 is restricted to the maximum of 500 AUD/MWh for a clearer presentation, the actual levels at 0.99 quantiles are around 93 AUD/MWh for the case without constraints and 1500 AUD/MWh for with constraints. It is clear that when conditioning on constraints, the levels of price at higher quantiles increases significantly, but the centre of the price distribution only changes a little. That is, with transmission constraints, the price remains centred on a level which is similar to that without constraints, but it becomes more volatile and increases significantly in the upper tail of the price distribution. Consequently the price risk involved in the trading activities when the network is constrained can differ

substantially from the one without constraints, which in turn serves to emphasise once again the importance of transmission constraints.

5.5 Conclusion

Accurate forecasts of electricity prices are important because they provide crucial information for market participants which, in turn, shapes their strategic decisions on hedging and trading activities. Despite significant advances in our understanding of deregulated electricity markets and in econometric methodology insofar as it pertains to modelling electricity prices, the covariates traditionally used to explain variations in electricity prices have remained relatively static. These covariates, such as load, temperature and reserve margins, are smoothly-varying leaving only the lagged price capable of capturing the rigidity and volatility of spot prices. In contrast, this chapter investigates the role of transmission constraints as a fundamental contributor to the variability of electricity prices. In order to identify the effects of transmission constraints, high frequency five-minute data made available by AEMO is employed. Although data on transmission constraints and bid profiles used in this study might not be accessible in some electricity markets, regional price information is widely available and consequently the differences between regional prices can be used as an indicator of transmission constraints.

The results reported in this chapter are based mainly on quantile regressions, which are used to ensure robust estimation and inference in the presence of extreme price outliers. It is found that transmission constraints contribute significantly both to the level and variability of price. Consequently, the performance of a price forecasting model is likely to be improved by incorporating information on transmission constraints. It is demonstrated that the presence of constraints is a necessary condition for the occurrence of extreme prices. One result of particular interest is that price is explosive in the upper tail of the price distribution - a result driven entirely by the presence of constraints. Clearly, there is need for further research to identify the other fundamental drivers of extreme prices given that transmission constraints are not sufficient conditions for these events.

To conclude, while the vast majority of existing studies treat the regional price process in a univariate framework, the importance of transmission constraints, as identified in this chapter, suggests that in interconnected markets it may be advantageous to consider treating regional markets as an integrated spatial system. More importantly, the important effects of transmission constraints on regional electricity prices identified in this chapter, combined with the real-time balanced market implies the possibility of large-scale base generators being in a position to exercise market power in areas that are frequently subject to transmission constraints. This leads to the subject matter of the next chapter.

Chapter 6

Strategic bidding and rebidding in electricity markets

6.1 Introduction

This chapter is concerned with an important policy conundrum faced by regulators in electricity markets world-wide, namely, the need to ensure the reliability of the supply while simultaneously promoting market efficiency through competition (Zhang, 2009; Bosco et al., 2012; Lízal and Tashpulatov, 2014; Bustos-Salvagno, 2015). This question is addressed in the institutional context of the Australian National Electricity Market (NEM) and particularly the regional market for the State of Queensland. In July 2007, the Queensland government started to implement a move towards full retail competition in the market. The operation of the NEM in Queensland changed significantly with the sale of two partly government-owned energy retailers and further reductions in government-owned generation capacity.[1] In another structural development of note, privately-owned electricity retailers embarked on a substantial program of investment in generation capacity mainly in the form of gas-fired turbines. These plants have a higher marginal cost than the base-load (mainly coal-fired) generators and therefore are used primarily as marginal (peaking) generators.

[1] State Government involvement in electricity generation in 2014 stands at about 50% down from around 65% in 2007 and there is no government presence in retailing.

Along with these structural changes, the market rules of the NEM were largely unchanged. The most significant exception was the large increase in the level of the market price cap prior to the introduction of competition and subsequent increases since then. These changes to the market cap price are plotted in Figure 6.1, showing the sharp increase from A$5,000/MWh to A$10,000/MWh in April 2005 and the subsequent increases to the current level of A$13,500/MWh. The reason for the successive changes in the price cap was to facilitate competition by giving participants more freedom to conduct their trading activities free from a binding price cap constraint. Another reason for having such a high cap price is due to the 'missing money' issue that is related to marginal generators (Simshauser, 2010b; Cepeda and Finon, 2011; Briggs and Kleit, 2013). That is, marginal generators in a market have the essential role of preventing black-outs in peak demand periods. Nonetheless, due to the high marginal cost and the limited peak demand periods, marginal generators are normally dispatched and paid only in the limited peak demand intervals. Thus the prices at these intervals need to be high enough for allowing marginal generators to cover both their fixed and marginal costs, and consequently to be able to exist. So in part, raising the market cap price of the NEM to the current level of A$13,100/MWh reflects this necessity. The fundamental question remains, however, of whether or not increases in the market cap actually had the (unintended) consequence of promoting strategic behaviour by market participants which in turn significantly affected electricity prices.

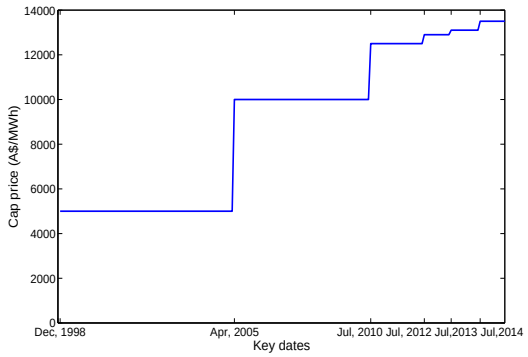

Fig. 6.1 The changes of market cap price in the NEM. Dec 1998 is the starting date of the NEM. In July 2012, the cap price starts to be adjusted annually according to the inflation.

A re-evaluation of the policy rules governing the NEM is overdue. Since the advent of deregulation almost 20 years ago, the rules of engagement have evolved and changed without being subject to rigorous examination (for studies of the NEM, see, Wolak, 2000; Outhred, 2000; Short and Swan, 2002; Hu et al., 2005; Tamaschke et al., 2005; Simshauser, 2010a). There is certainly enough casual empirical evidence to support the claim that prices in the NEM have become increasingly volatile in recent years. The crucial question of course is whether this volatility is linked to strategic behaviour and therefore whether current policy with respect to regulation of the electricity market needs to be revisited. To address this question, a detailed dataset based on 'DispatchIS Reports', 'Yesterdays Bids Reports' and 'Next Day Dispatch', all of which is publicly available from the market operator is built.[2] Data for dispatch prices, dispatched quantities, loads, inter-regional flows and bids and actual dispatch of every generator in the market are recorded at a five-minute frequency for the period from 04:05, 30 November 2008 to 04:00, 24 June 2014, a total of 681,696 observations, is used to explore the bidding behaviour of market participants.

The central findings of the chapter are that interesting irregularities in the movement of price, inter-regional flow, load and dispatched capacity, can be related directly to the interaction of the physical characteristics of electricity markets and the strategic behaviour of market participants. These irregularities are exacerbated by the current policy rules governing bidding behaviour in the NEM. Policy options for dealing with these problems are also discussed.

6.2 Institutional framework

The Australian National Electricity Market (NEM) is one of the world's largest deregulated electricity markets which comprises the regions of New South Wales (NSW), Victoria (VIC), Queensland (QLD), South Australia (SA) and Tasmania (TAS). These five regional markets, are further subdivided into smaller areas called nodes. A transmission network that links the generators with load centres covers around 4,500 km while the distribution network,

[2] See www.nemweb.com.au.

which transports electricity from points along the transmission network to final users, is 17 times larger at around 750,000 km. A feature of the NEM is that in eastern and southern Australia the different regions are fully interconnected. QLD and NSW are connected by two interconnectors, (QNI and Terranora) with total capacity of around 400 Mega Watts (MW) into QLD and 1,100 MW into NSW; VIC and NSW are connected by one interconnector with capacity of around 1,900 MW into VIC and 3,200 MW into NSW; Murraylink with capacity around 200 MW in both direction connects VIC and SA; and Basslink (around 500 MW) connects VIC and TAS. Based on this infrastructure, the trading between around 300 registered generators with total installed capacity of around 48,000 MW and nine million customers are operated as a pooled market under the supervision of the Australian Energy Market Operator (AEMO).

Some summary statistics for the dispatch price and load in various regions of the NEM from 1 June 2014 to 31 May 2015 are presented in Table 6.1. Based on the average load, it can be seen that the regions vary quite markedly in size. The standard deviation of load in each region is about 15% of the median load, indicating a moderate variation in the demand side of regional markets. In contrast, the median price for each region varies around the level of A$45/MWh, with the standard deviation being much higher in four of the five regions. The exception is NSW, which does not have such a high price variation because of its unique geographical location which affords it more interregional transmission capacity than the other states. The minimum and maximum prices reflect the existing market floor (A$-1,000/MWh) and cap (A$13,500/MWh) prices set by AEMO.

The dispatch prices shown in Table 6.1 are the outcome of the dispatch process. For each region, bids of quantity and price from all generators in the market are aggregated and a central dispatch algorithm then dispatches the available bids based on merit order (according to bid price), thus leading to the demand at each location generally being met by the available generators with the lowest bid price. A dispatch price is calculated every five minutes (a dispatch interval) based on the last generator dispatched (marginal generator) with the highest dispatched bid price. In each half-hour trading interval, the trading price for a region is quoted as the average of the six regional dispatch prices in that trading interval. The regional

Table 6.1 Summary information for the regions in the NEM, from 1 June 2014 to 31 May 2015. The price information is based on regional dispatch prices.

	Queensland	New South Wales	Victoria	South Australia	Tasmania
Min. price	-1,000	-1,000	-1,000	-1,000	-599
Median price	43	46	43	42	41
Max. price	13,500	349	13,500	13,500	13,500
S.D. of price	107	19	174	143	79
Min. load	4,073	5,050	3,585	858	678
Median load	5,885	8,315	5,630	1,447	1,089
Max. load	8,757	12,217	9,386	2,955	1,613
S.D. of load	820	1,253	857	289	135

trading price is used as the settlement price for all transactions in that trading interval and all dispatched generators in the region are paid by this uniform regional price irrespective of their bid prices. This price setting process is known as locational pricing (Ding and Fuller, 2005). It has the advantage of allowing the price to be based on regional specific equilibrium conditions, thus encouraging competition in regional specific markets.

A feature of the electricity market is the important role of the transmission grid in an integrated electricity market because all locations within a region must be in equilibrium. Any transmission constraints on the grid can lead to the isolation of some areas within the region and a mismatch of local supply and demand. Consequently, in order to achieve a highly adaptive supply side which can respond effectively to the real-time market changes, generators in the NEM are allowed to adjust their bids at any time up to five minutes before the actual dispatch.[3]

Flexible bidding allows the supply side to adjust quickly and thus helps ensure the maintenance of real-time equilibrium. The bidding process also gives market participants a high degree of freedom in conducting their trading strategies, which theoretically should promote competition in the market. However, as will become apparent, flexible bidding

[3]Recall the bidding rules introduced in Section 1.2. A generator needs to submit its bid day ahead with the offered capacities associated with ten price bands that are located in the range between the floor and cap prices. Adjustment on the offer capacity in each price band is allowed five minutes before the actual dispatch. For more details, see, Australian Energy Market Commission (2015).

may also enable generators to bid strategically with the undesirable consequence of enabling regional prices to deviate from the fundamental cost of generation.

6.3 The anatomy of price spikes

An important feature of deregulated electricity markets worldwide is the intermittent occurrence of abnormally high prices (price spikes) in spot electricity markets (Barlow, 2002; Escribano et al., 2002; Lucia and Schwartz, 2002; de Jong and Huisman, 2003; Byström, 2005; Cartea and Figueroa, 2005). Both the size of these irregular price events and their duration are potentially harmful to electricity retailers who have to manage the associated price risk (Anderson et al., 2007). Price spikes occur when the price of electricity exceeds a given price threshold. The early literature in the Australian market took the threshold to be a constant value usually in the region of A$80 - 100/MWh, based on the fact that the dispatch price fluctuates between A$30 and A$70/MWh in normal conditions (Christensen et al., 2009, 2012). The identification of the threshold for defining a price spike can also be variable and model dependent, see, for example, Cartea et al. (2009), Janczura et al. (2013) and Zareipour et al. (2011). Weron (2014) provides a comprehensive recent survey of modelling prices and price spikes in general.

To give a feel for the volatility of the dispatch price, a time series plot of the Queensland dispatch price is shown in panel (a) of Figure 6.2. Prices at both the market cap price and the market floor price are observed in many instances. Either positive or negative extreme prices tend to revert back to a normal level very quickly and in many instances last only for a five-minute dispatch period. Moreover, extreme price events tend to cluster around the beginning of every year when it is during the summer season in Queensland. Panel (b) of Figure 6.2 shows a scatter plot of dispatch price against dispatched quantity, which illustrates the large disparity in prices which occurs at the similar level of dispatched quantity. Put differently, abnormally high dispatch prices occur at a wide range of dispatched quantities and are not merely limited to situations which the system can reasonably be characterised as being at capacity.

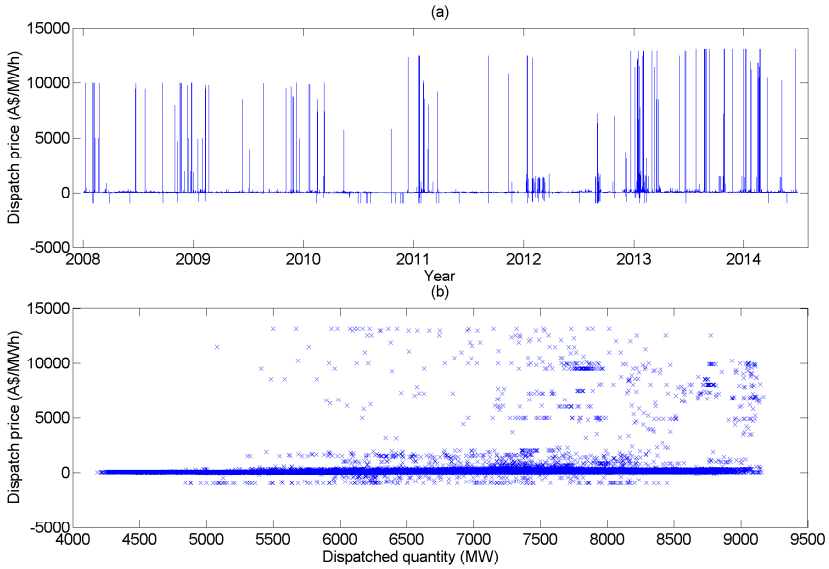

Fig. 6.2 Panel (a), a times series plot of dispatch prices. Panel (b), a scatter plot of dispatch prices and dispatched quantities for Queensland, from 04:05, 30 November 2008 to 04:00, 24 June 2014.

The standard explanation for the occurrence of abnormal price events is a micro-theoretic one. If demand rises to the point of system capacity indicated by the vertical portion of the supply curve, due perhaps to extreme weather conditions, then even a small change in demand will have a significant effect on price. Similarly, if a significant portion of base supply suddenly goes offline due to generation failure, then the vertical portion of the curve will be shifted closer to the origin and the existing normal level of demand may be sufficient to cause a significant price increase. In other words, under this traditional explanation, episodes of price spikes are simply a manifestation of scarcity and are not due to strategic behaviour on the part of market participants.

There is some casual empirical evidence to support this argument. Consider Figure 6.3 which shows the 288 five-minute bid curves on 28 August 2013, in which the large changes in the bid quantities at the lower price levels might on the surface suggest there is a sudden loss of large generation capacity. Consistent with this view is the fact that the very steep section of the bid curve changes position over the course of the day and region of variation is very close to 2,000 MW.

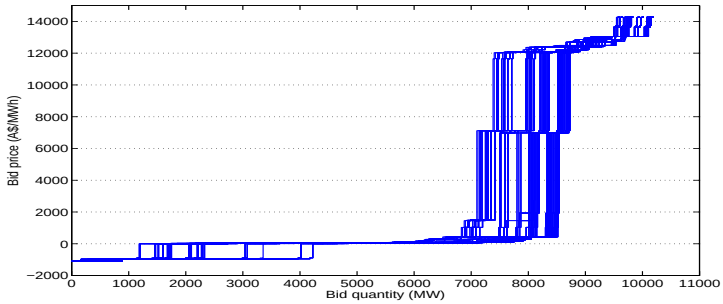

Fig. 6.3 The 288 five-minute bid curves for Queensland, 28 August 2013.

This evidence is however not completely consistent with the argument. At the market cap price the total available capacity in that day only varies around 500 MW, which is in sharp contrast to the variation of around 3,000 MW at the market floor price. If the shift in the bid curves at the floor price level is indeed due to the failure of a large base generator, the same shift would also occur at the maximum price level so that the loss of generation capacity is reflected in total available capacity. Furthermore merely observing shifts in the bid curves is not informative unless combined with the timing of the price spike.

The scatter plot of load against the logarithm of the half-hourly trading price in Figure 6.4 provides another way of visualising the differences that can arise in the pattern of abnormal price events. In June 2007 Queensland was in the grip of a severe drought which limited the cooling water available for coal- and gas-fired base load generators and also reduced the amount of water available for hydro generation. The top panel of Figure 6.4 therefore shows a situation in which the system capacity is constrained due to generation failure. Here price is regularly above the threshold price at all levels of load.

The bottom panel of Figure 6.4, the months of January and February 2009, tells a different story. Here abnormally high prices occur mainly at high levels of load. The fact that these extreme price events primarily occur at high loads, *but not solely at system capacity*, suggests that many of these irregular events may be attributable reasons other than the depletion of cheaper generation capacity.

Focusing on one extreme price event on 28 August 2013 which occurred at a moderate load level, Table 6.2 shows a typical half-hour trading interval which contains an abnormally

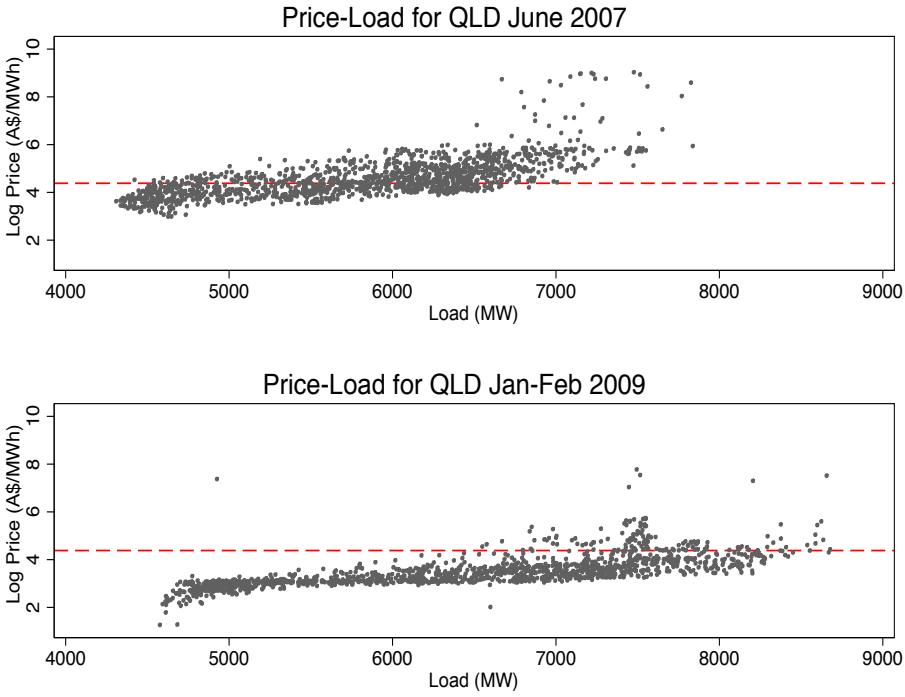

Fig. 6.4 Scatter plots of system load versus the logarithm of price for Queensland in the months of June 2007 (top panel) and January and February 2009 (bottom panel). The dotted line represents the threshold value (natural logarithm of A$80/MWh) above which an irregular price event occurs.

high price spike at low dispatched system capacity. This particularly intense price spike, which resulted in the (then) market cap price of A$13,100/MWh being reached, occurs at 06:40 but the dispatch prices in the five minute intervals immediately before and after the spike are at normal levels. Note also that while the load (regional consumption) is stable at around 5,600 MW for the entire period, the price drops after the spike are accompanied by significant **increases** in dispatched quantity and inter-regional outflow (over 1,000 MW respectively).[4]

To summarise, the discussion given in this section gives rise to three important questions. *First*, why is it that an extreme price event can occur even at low levels of load? *Second*, why does the dispatch price exhibit such extreme behaviour over a very short period? *Third*,

[4]By convention a positive flow indicates that Queensland is exporting electricity to the adjacent region of New South Wales.

Table 6.2 Dispatch prices, load, dispatched quantities and inter-regional out-flows of Queensland in a typical half-hour interval with a price spike.

Time (28 Aug 2013)	Dispatch prices (A$/MWh)	Load (MW)	Dispatched quantities (MW)	Inter-regional out-flow (MW)
06:35	66	5,595	5,458	-137
06:40	13100	5662	5,500	-161
06:45	50	5,524	5,993	469
06:50	46	5,603	6,476	873
06:55	29	5,589	6,687	110
07:00	47	5,770	6,889	874

why does the dispatch price drop significantly after a spike despite the increase in actual generation together with an increase in regional outflow?

6.4 Strategic bidding

Attempts by generators to manipulate the dispatch price of electricity by withholding capacity is now a well documented phenomenon (Bunn and Martoccia, 2005; Biggar, 2011; Karthikeyan et al., 2013; Schwenen, 2014). Essentially the idea is that generators in deregulated markets have an incentive to behave strategically. By reducing the capacity offered at their cost of generation they can ensure that the equilibrium price (determined by the bid of the marginal generator) is higher than it otherwise would be. Recent evidence on capacity withholding includes Nappu et al. (2013) who examine the relationship between capacity withholding and transmission constraints in a hypothetical market; Maenhoudt and Deconinck (2014) who investigate the profit maximising strategy of generators in a simulated market and in the Iberian electricity market; Zhang et al. (2015) look at renewable energy generators and capacity withholding in the PJM market in the United States; and Koschker and Möst (2015) also examine capacity withholding in the context of the German electricity market. These papers are largely empirical in nature and their theoretical framework is simply understood to be that capacity withholding arises as a natural consequence of the exercise of market power by individual generators on the supply side of the market.

In the NEM, capacity withholding arises as a consequence of the market clearing arrangements. Because the half-hourly settlement price is the average of each of the six five-minute dispatch prices, within each half-hour interval there exists an incentive on the part of the generators to withhold capacity. Clearly, if an abnormally high dispatch price is recorded for *any* of the six five-minute dispatch intervals, then the average settlement price received over the half hour will be forced well above the marginal cost of electricity generation.

The concept of strategic bidding, as used in this paper, is a generalisation of capacity withholding and involves both capacity withholding at low prices (representative of the cost of generation) and at the same time bidding all remaining capacity at the maximum allowable price. The profitability of bidding excess capacity at the maximum price allowed by the market operator stems from two key technical constraints faced by generators, namely their *ramp-up* rates and their *synchronisation* times. The ramp-up rate of a generator is the speed of increase in generation capacity for an *in-use generator* that has not yet reached it's maximum capacity, while the synchronisation time refers to the minimum time taken by an *idle generator* from receiving a dispatch target to the delivery of the target to the grid. It stands to reason that the ramp-up rate of a generator that is already producing electricity is generally a lot faster than the synchronisation time for any type of generator that is currently not dispatched[5]. For example, a generator that is already generating electricity but not at capacity, can increase its output by potentially more than 200 MW per minute, while idle generators take between 5 and 30 minutes to come online[6]

There is a practical implication of the fact that synchronisation times are generally much lower than the time required for a generator to ramp-up its output. Generators which are already dispatched in previous intervals, but not at full capacity, are really the only choice in terms of fulfilling additional load, should the need arise, in the next dispatch interval. This gives them a strategic advantage over the marginal generators that were not previously

[5]For detailed descriptions, see http://www.aemo.com.au/Electricity/Market-Operations/Dispatch/Fast-Start-Inflexibility-Profile-Process-Description and Participant Input Interface Energy-MNSP-FCAS Bid File Submission available at http://www.aemo.com.au/About-the-Industry/Information-Systems/Using-Energy-Market-Information-Systems

[6]Oil generators generally have the fastest synchronisation time while coal and gas generators are a lot slower.

dispatched. The optimal bidding strategy for a hypothetical base generator operating in the NEM, could quite conceivably be as follows:

(i) bid a reduced capacity at lower price level to ensure continuous dispatch; and

(ii) bid the remaining capacity at the maximum allowable market price.

This strategy is referred to as bid splitting and it enables the base generators to drive up the price by reducing capacity at the lower price levels while ensuring that they received the final settlement price by being continuously dispatched over the half-hour interval.

To examine the empirical evidence concerning this hypothetical bid strategy consider Figure 6.5 which shows the detailed breakdown of the bid curves before and at a price spike. The intersections of dashed lines indicate the actual dispatched quantities and dispatch prices. To provide a clearer illustration of the differences in bidding strategy, the generators' bids in Figure 6.5 are decomposed into three groups. The generators represented by squares in Panels (a) and (b) make bids which span over a large range of prices. These generators are predominantly coal-fired, base-load generators and collectively they contribute about a third of the total generation capacity in Queensland. There are a considerable number of bids at the very extremes of the price range, but there is little or no capacity offered in the intermediate price range between A$1,000 and A$11,000/MWh. Panels (c) and (d) represent with circles the bids from generators who only have bids placed over the pre-spike price level of A$66.03/MWh. Finally, the crosses shown in panels (e) and (f) represent the bids from the remaining generators who are only concerned with being dispatched and accordingly bid all their capacity at the lowest prices.

Figure 6.5 provides clear evidence of bid-splitting behaviour by the large coal-fired generators. It is absolutely clear that these generators bid substantially less than their full capacity at lower prices, to ensure that they will be dispatched, while simultaneously bidding capacity at the market cap price. Hypothetically, the bids marked in circles in Panels (c) and (d) should play an important role in providing peaking capacity at reasonable prices should the need arise. It is here that the crucial importance of ramp-up versus synchronisation is clearly seen. Because these marginal generators only bid at price levels above A$66.03/MWh

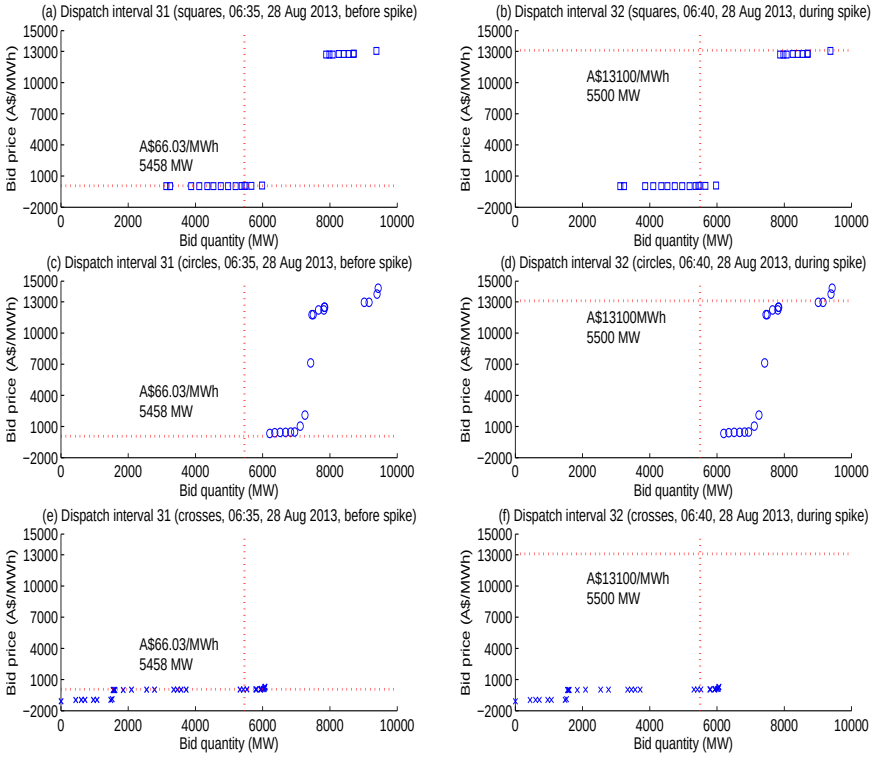

Fig. 6.5 Detailed bid curves in two consecutive dispatch intervals. Panel (a), (c) and (e) are the bid curve at 06:35, 28 Aug 2013 (before a price spike). Panel (b), (d) and (f) are the bid curve at 06:40 28 Aug 2013 at the time of a price spike. The dispatch prices and dispatched quantities are denoted by dotted horizontal and vertical lines with the values indicated beside. The bid curves are shown for three separate groups. The square boxes shown in Panel (a) and (b) are bids from generators with wide bid price range (with generator ID, generation capacity and fuel type: TNPS1 (443 MW, coal), TARONG#1 (365 MW, coal), TARONG#3 (365 MW, coal), STAN-1 (365 MW, coal), STAN-2 (365 MW, coal), STAN-3 (365 MW, coal), STAN-4 (365 MW, coal), GSTONE4 (280 MW, coal) and SWAN_E (370 MW, gas)). The circles shown in Panel (c) and (d) are bids from generators that only bid at the level above A$66.03/MWh. The remaining bids are shown as crosses in Panel (e) and (f). All bids are arranged in merit order.

(above the pre-spike price level) they are not dispatched prior to the spike. Consequently, when the need arises to dispatch more load, the speed at which existing base-load generators can ramp-up their generation means that at the time of the spike one of the in-use base generators using a bid splitting strategy (indicated by a square) is dispatched. The peaking generators are not dispatched simply because transmission constraints or the time it takes for them to synchronise is longer than the required five minute interval. Consequently, the regional price jumps from A$66.03 to A$13,100/MWh and the lower price bids of the generators represented by circles and crosses are not dispatched.

It is therefore evident that the price and bidding behaviour of large scale base generators before and during an extreme price spike is consistent with the bid splitting hypothesis. It is also clear that being able to rely on ramping-up output as opposed to becoming synchronised within a five-minute interval, confers a strategic advantage on the base generators. However, it could still be argued that bid splitting is simply a legitimate (as opposed to a strategic) way of avoiding dispatch, due perhaps to technical reasons such as fuel shortages, maintenance, or temporary outages in generation capacity. This point of view may be refuted on two grounds. The first is apparent in Panels (e) and (f) of Figure 6.5 where quite standard bidding behaviour by generators results in a certain amount of capacity remaining idle. It follows therefore that avoiding dispatch can be accomplished by quite standard bidding patterns. The second way of refuting this argument requires examination of the rebidding behaviour of the large generators after the occurrence of a price spike.

The empirical analysis of this section so far focused on data at the aggregate level, namely, changes in bids at the level of the regional market. In the remainder of this section, attention is shifted to individual generators in order to obtain statistical evidence of differences in bidding behaviour. One of the consequences of bid splitting is an increased span for bid prices (of the order of A$10,000/MWh), a result which obviously cannot be justified by marginal generation costs. To see how the span of bid price changes for different generators, the bids from three generators, two coal fired plants Stanwell unit 3 (STAN-3) and Kogan Creek (KPP_1) and one diesel plant, Mackay GT (MACKAYGT) are considered. A summary of the characteristics of these three generators is listed in Table 6.3.

Let the difference between the lowest and the highest bid prices at dispatch interval t with positive generation offered be denoted $\Delta\vec{B}_t$. Recall that a trading interval is a half-hour period divided into six dispatch intervals. Consider the quantile regression

$$\Delta\vec{B}_t = \beta_1(\tau) + \beta_2(\tau)D_{0t} + \beta_3(\tau)D_{1t} + \beta_4(\tau)D_{2t} + \beta_5(\tau)D_{Rt} + \beta_6(\tau)N_t + u_t, \qquad (6.1)$$

Table 6.3 Summary of the characteristics of three generators considered in the comparison. The sample period is from 1 Jan 2008 to 24 Jun 2014 (681,696 of total dispatch intervals). The number of dispatched intervals for each generator is calculated as the number of intervals where the dispatched price exceeded the minimum bid price (with positive available capacity) of the generator.

Generators' ID	STAN-3	KPP_1	MACKAYGT
Fuel type	Coal	Coal	Diesel
Maximum capacity (MW)	365	740	34
Number of dispatched intervals	645,394	593,893	720
Number of intervals with available capacity	645,762	594,254	336,832
Total dispatch intervals		681,696	

where τ represents the relevant quantile taken to be 0.01 to 0.99 in 50 equally spaced steps. The variable D_{0t} takes the value 1 if a price spike of greater than A$ 5,600/MWh is observed.[7] The variable D_{1t} represents the first dispatch interval after a spike and takes the value 1 if $D_{0t-1} = 1$. The variable D_{2t} is similarly defined and refers to the second interval after the spike, but only takes the value 1 if it falls in the same half hour trading interval as the price spike. The variable D_{Rt} mops up all the remaining dispatch intervals in the current trading period. The variable N_t takes the value 1 in the first dispatch interval of a new half hour trading period if either $D_{2t-1} = 1$ or $D_{Rt-1} = 1$.

To obtain the coefficient estimates, the Bayesian quantile regression approach of Yu and Moyeed (2001) is used with the Gibbs sampler of Kozumi and Kobayashi (2011). More specifically, in this approach the residual term u_t is assumed to follow a scaled asymmetric Laplace distribution. By writing the scaled asymmetric Laplace distributed u_t as a mixture of a standard exponential variable and a standard normal variable, it becomes straightforward to implement the Gibbs sampler for u_t. Following Kozumi and Kobayashi (2011), the priors for all parameters are assumed to be normally distributed with mean 0 and large variance. Finally, posterior inference is based on 10,000 draws with the first 5,000 discarded as burn-in period.

[7] The value of A$5,600/MWh is chosen because if a price spike equal to A$5,600/MWh occurs in the first dispatch interval and all other dispatch prices then settle at the floor price, the averaged trading price for the interval is A$100/MWh, which represents a solid return above the marginal cost of base generators.

Obviously, for each quantile of interest τ, differences in the coefficient estimates for the three generators will reveal different bidding strategies. However, there are two coefficients of particular importance. The coefficient $\beta_1(\tau)$ represents the level of the bid span at (or below) which the generator operates for $\tau \times 100$ % of the time. If a high bid span is representative only of temporary outages, then this coefficient should take high values only for the very high quantiles. The coefficient $\beta_6(\tau)$ is an offset for $\beta_1(\tau)$ and indicates the degree with which the level of the bid span adjusts if there was a price spike in the previous trading interval.

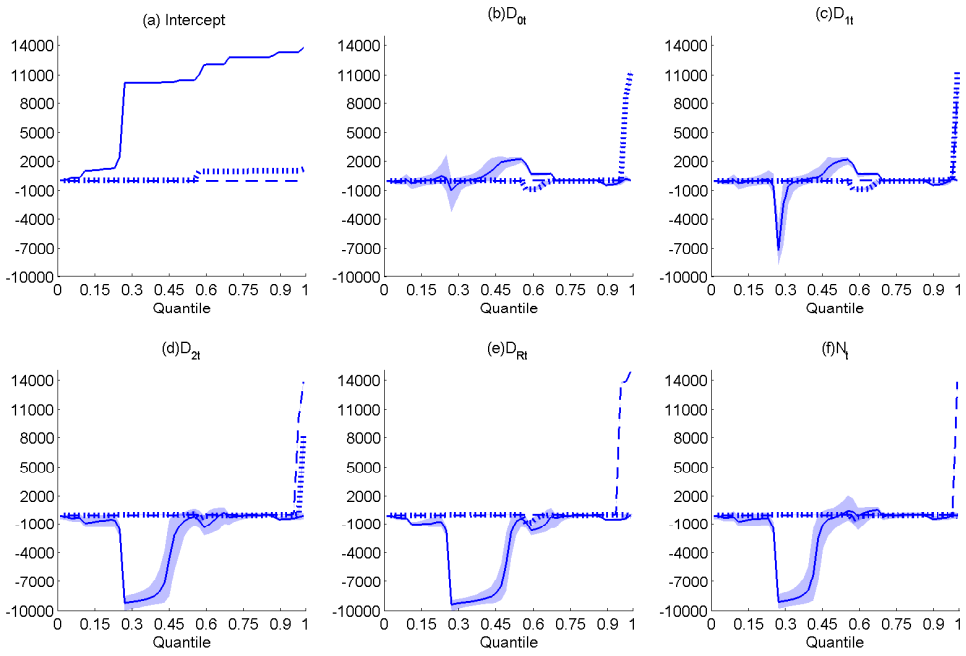

Fig. 6.6 Quantile regressions for the spans of bid prices of generators with ID STAN-3 (solid line), KPP_1 (dotted line) and MACKAYGT (dashed line). 95% credible intervals are shown as shaded areas. From Panels (a) to (f), the estimates for the parameter of constant and dummy variables defined after Equation (6.1) based on a grid of 50 equally spaced quantile from 0.01 to 0.99.

The differences in the bid patterns of the three generators are illustrated in Figure 6.6. It can been seen in Panel (a), the intercept, $\beta_1(\tau)$ for STAN-3 (solid line) is over A$10,000/MWh for around two thirds of the total dispatch intervals. This is completely different from the pattern of the intercepts recorded for KPP_1 (dotted line) and MACKAYGT (dashed line). The result holds particular significance in terms of the differences between

STAN-3 and KPP_1, which are similar base load generators with the same fuel type. This result is clearly indicative of strategic bidding behaviour on the part of STAN-3 as the behaviour of the intercept is obviously not due to temporary outages of limited duration.

Another interesting result illustrated in Panels (d) and (e) of Figure 6.6 relates to coefficients $\beta_4(\tau)$, the coefficient on dummy variable for the second dispatch interval after a spike (D_{2t}), and $\beta_5(\tau)$, the coefficient on the dummy variable for all the remaining dispatch intervals in the same trading period as the spike (D_{Rt}). For the generator STAN-3, these coefficient estimates become strongly negative at about the 0.3 quantile. This indicates that there is a significant change in bidding behaviour two periods after a spike with the span of the bid price being significantly reduced. This observed behaviour is different from the almost constant estimates of $\beta_4(\tau)$ and $\beta_5(\tau)$ for the other two generators, KPP_1 and MACKAYGT. In fact, this pattern is a manifestation of another strategic aspect of generator bidding behaviour known as rebidding, which is the subject matter of the next section.

6.5 Rebidding

Recall from the market rules introduced in Section 6.2 that generators are allowed to change their bids up to five minutes before actual dispatch. Also, generators are paid by the half-hour average regional trading price rather than the five-minute dispatch price. Given these rules, the obvious profit maximising strategy for a base-load generator *after the occurrence of a price spike* is to rebid all available generation capacity at the lowest allowable price. Consider, for example, that a price spike (of maximum allowable size) occurs in the first five-minute dispatch period of a half-hour trading interval and the dispatch price then drops to the floor price (A$-1,000/MWh) for the remaining five dispatch periods. The actual settlement price for that half hour based on averaging will be about A$1333/MWh. This price is well above the marginal cost of any type of generator (A$40 to A$60/MWh) and therefore any risk associated with this rebidding is eliminated.

Very strong evidence of rebidding after a price spike is found in Figure 6.7. Panels (a) and (b) are the total bid curves which are aggregated versions of the bid information illustrated

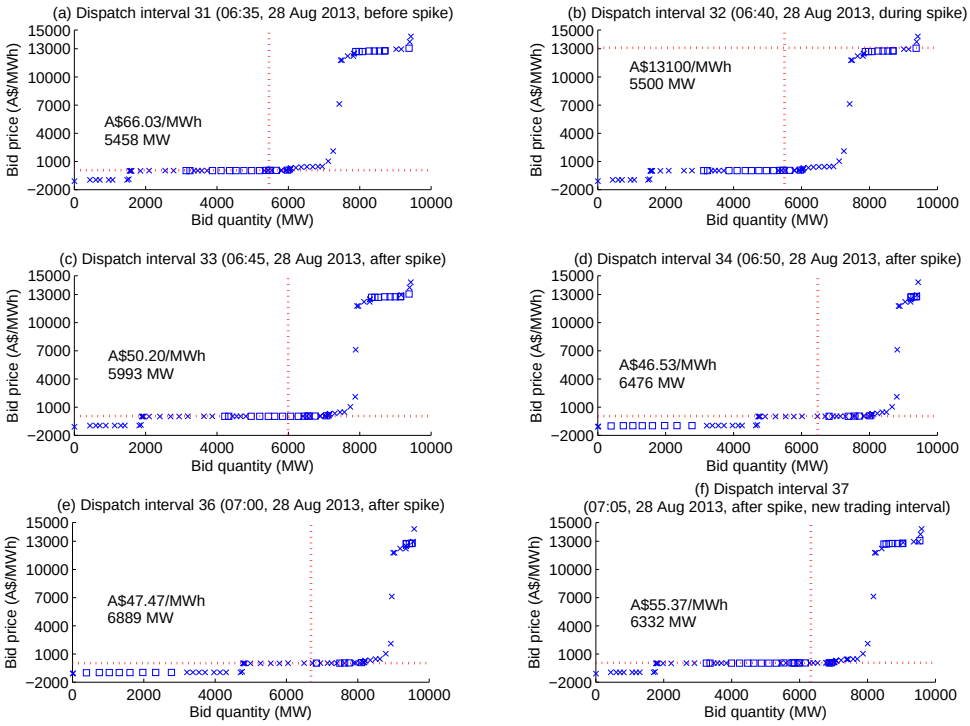

Fig. 6.7 Illustrating the rebidding behaviour of base-load generators in Queensland between 06:30 and 07:00 on 28 August 2013.

in Figure 6.5. Panel (c) of Figure 6.7 shows the first dispatch interval after the spike, in which the dispatched quantity increases by nearly 500 MW and the dispatch price drops back to a normal level without any significant change in the bid curve. These effects are due to the cheaper marginal generators becoming synchronised. Panel (d) shows the second dispatch interval after the spike in which the price drops further despite another nearly 500 MW increase in dispatched quantity. In this interval, the majority of boxes representing the bids from large scale in-use base generators that used a bid splitting strategy have now shifted to the market floor price. Recall the market rules introduced in Section 6.2 that a rebid can be made, at the latest five minutes before dispatch. Assuming that generators observe the price spike first and then rebid as soon as possible, the second dispatch interval after the spike is the earliest interval for the rebids to become effective. Moreover, due to the uniform regional price, the large scale rebidding from the second dispatch interval onward is a region wide event.

To provide further evidence of the rebidding behaviour and see whether it can be generalised to other trading intervals with price spikes, define ΔB_t as the change in total available bid quantity in the regional market at the current dispatch price at time t. To obtain the value of this variable, the bids from all generators in Queensland at a five-minute frequency are cumulated to obtain the five-minute bid curve for Queensland. By substituting the dispatch price at t into the bid curve at both time t and $t-1$, the two available bid quantities at the current price level are obtained. Then ΔB_t is set to the available bid quantity at t minus the available bid quantity at $t-1$. Consequently, ΔB_t represents the bid change at the current and most relevant price level. A time series plot of ΔB_t is shown in Panel (a) of Figure 6.8 and a histogram of ΔB_t is shown in Panel (b). In general, it may be concluded that ΔB_t varies moderately most of the time with only a limited number of extreme values and is concentrated around 0 with an almost symmetric distribution.

To focus only on the observations of ΔB_t in the trading intervals that are relevant to the rebidding described, consider only those trading intervals with at least one price spike that is over A$5,600/MWh and where this spike only occurs in the Queensland region. The restriction of a Queensland only price spike allows a clearer identification of rebidding because spikes occurring in multiple regions concurrently generally indicate an exhaustion of generation capacity at peak demand intervals. Given the whole sample period from 04:05 30 November 2008 to 04:00 24 June 2014, there are 155 observations of ΔB_t shown in Panels (c) and (d) of Figure 6.8 that satisfy these conditions. It can be seen that in Panel (c), when spikes occur, the changes in bids are mostly negative. Although this may be caused by outages of generators instead of strategic bidding, this pattern is highly suggestive of capacity withholding. In Panel (d), the change in bid quantities becomes largely positive. Since this is the first dispatch interval after the spikes, the bid changes are the results of rebidding before observing the spikes. The large number of positive changes in bid quantities means that generators are either pre-empting the occurrence of a price spike or frequently shifting their bids to a higher price and then moving back to a normal price level in order to create spikes.

If a price spike occurs in the last two dispatch intervals of a half-hourly settlement period, then generators do not have enough time to respond given rebidding requires that a price

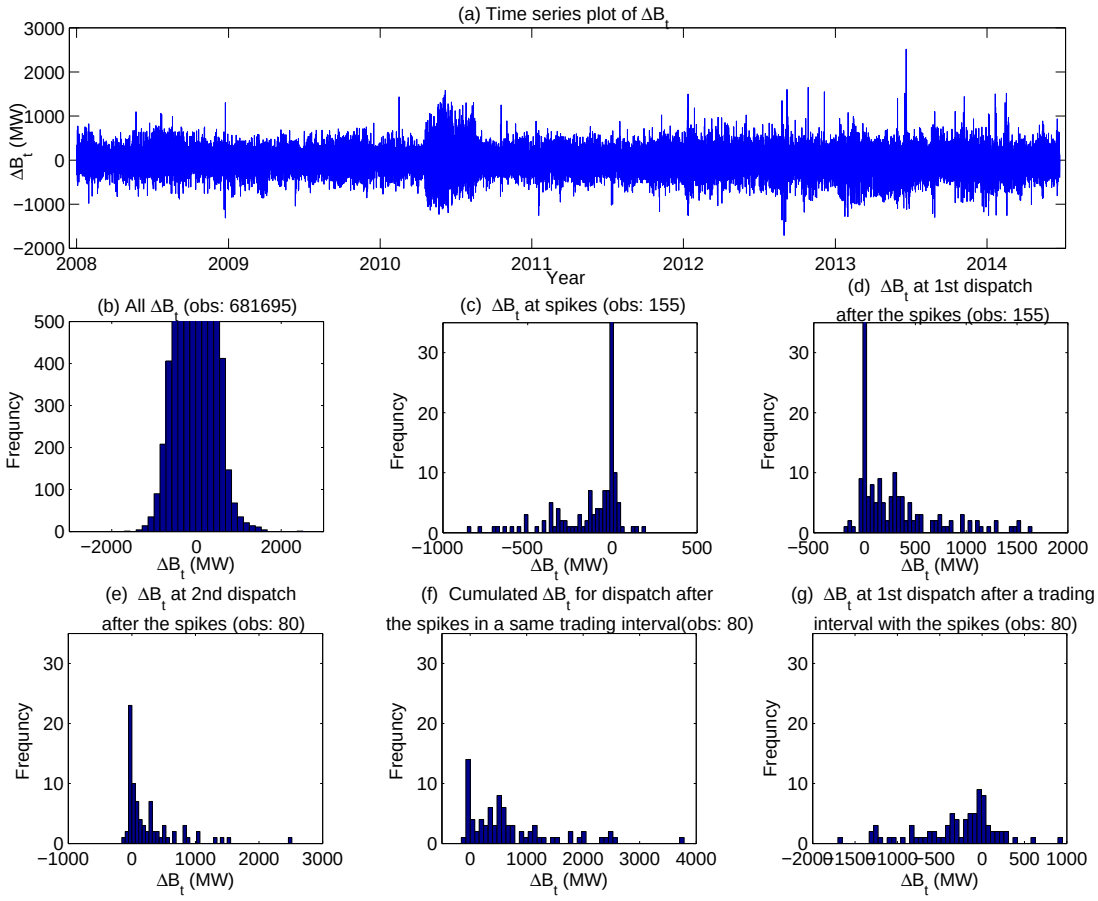

Fig. 6.8 Panel (a), time series plot of changes in bid quantity (ΔB_t). Panel (b), histogram of ΔB_t for all t. Panel (c), histogram for ΔB_t in the dispatched interval with a price spike above A$5,600/MWh. Panel (d) and (e), histogram of ΔB_t for the first and second dispatch intervals of a trading interval after a price spike (over A$5,600/MWh) was occurred in the one of the first four dispatch intervals of that trading interval. Panel (f), the cumulated changes in bids in the remaining dispatch starting from the second dispatch intervals after the spike to the last in the same trading interval. Panel (g), histogram of ΔB_t in the first dispatch interval after the trading intervals with the spikes.

spike be observed first. Accordingly Panels (e), (f) and (g) of Figure 6.8 are limited to those 80 observations of ΔB_t which follow immediately after the 155 observations displayed in Panel (d) and where at least one price spike occurred in any of the first four dispatch intervals. It is apparent that starting from the second dispatch interval after a price spike, Panel (e), the overwhelmingly positive changes in bids are convincing evidence of the rebidding behaviour discussed. Furthermore, note that the largest quantity of ΔB_t is over 2,500 MW which is over one third of the usual dispatched quantity in Queensland! In the remaining dispatch intervals

after the first price spike, the cumulated changes in bid quantities are almost all positive as shown in Panel (f), which is also in accordance with the rebidding phenomenon. Finally, in a new trading interval after the spikes, Panel (g), negatively skewed ΔB_t shows that the bids start to move back to higher price level because of the expected drop in trading price in a new trading interval, which again aligns with the rationale of rebidding.

The information shown in Figure 6.8 is strongly suggestive of generator rebidding after price spikes. To provide statistical evidence of the rebidding pattern shown in Figure 6.8, similar quantile regressions to those used in Section 6.4 are implemented. The same linear functional form shown in Equation (6.1) is used with identical dummy variables. The difference is that the dependent variable in this section is now taken to be ΔB_t, the change in total bid quantity for the regional market. Consequently, the main goal of the quantile regressions is to show the distributional features of the changes of bid quantity (ΔB_t) at trading intervals where potential rebidding is to be expected.

The definitions of the dummy variables mean that these variables capture (approximately) the information shown in Panels from (b) to (g) of Figure 6.8. The estimated coefficients at each quantile therefore indicate the levels of ΔB_t in different conditions represented by the dummy variables and at the quantile of τ, and highlight the differences in the distribution of ΔB_t. Specifically, in the presence of rebidding, the estimated coefficients for $\beta_{4\,\tau}$ and $\beta_{5\,\tau}$ should be significantly greater than 0 for most of the quantiles, whereas $\beta_{6\,\tau}$ should be significantly lower than 0 for most quantiles. Once again, the coefficient estimates are obtained by Bayesian quantile regression and all the results are based on 5,000 draws after a 5,000 of initial burn-in period. The quantiles are taken to be 0.05, 0.2, 0.35 0.5, 0.65, 0.8 and 0.95 and the resulting coefficient estimates, together with 95% credible intervals, are shown in Table 6.4,

The results reported in Table 6.4 provide a solid statistical confirmation of the information shown in Figure 6.8. In general, the change in bids (row 1) are highly concentrated around 0 in normal market conditions. On the other hand, the changes become positively skewed in the intervals after a price spike due to rebidding. Moreover, the highly negatively skewed changes in the last row also show that generators start to move their bids back to a higher

Table 6.4 Quantiles ($\tau = 0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95$) estimates of the distributions of ΔB_t in the dispatched intervals of all, around price spikes and of trading intervals with potential rebidding. (95% credible intervals based on the posterior draws are reported in brackets)

	τ =0.05	τ =0.2	τ =0.35	τ =0.5	τ =0.65	τ =0.8	τ =0.95
Intercept	−44 (−45,−44)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	44 (44,45)
D_{0t}	−472 (−496,−454)	−192 (−210,−177)	−72 (−84,−58)	−7 (−12,−2)	0 (−2,0)	0 (0,2)	−24 (−33,−11)
D_{1t}	10 (−2,21)	0 (−1,2)	53 (42,63)	175 (159,194)	311 (297,324)	518 (501,541)	1147 (1100,1185)
D_{2t}	7 (−5,21)	0 (−3,3)	9 (1,19)	99 (88,109)	256 (230,279)	478 (439,512)	1148 (1011,1259)
D_{Rt}	−2 (−20,13)	0 (−3,0)	0 (0,1)	5 (0,9)	57 (48,69)	151 (143,161)	544 (509,582)
N_t	−1203 (−1229,−1184)	−668 (−703,−634)	−369 (−381,−352)	−159 (−178,−141)	−42 (−64,−20)	9 (−1,30)	206 (181,231)

price level in the new trading interval. Based on the observations made in Figures 6.7, 6.8 and Table 6.4, the phenomenon of rebidding by generators after price spikes is conclusively established.

Rebidding generally leads to a large amount of capacity offered by multiple generators, at or near the market cap price in the dispatch intervals before a price spike, shifted to the floor price in the remaining dispatch intervals. These large scale shifts in supply have a particularly worrying consequence, namely, counter-price flow. Counter-price flow refers to the situation where even though the trading price in one region is high, the region is also exporting electricity to adjacent regions which have lower trading prices. To understand the association between counter-price flows and rebidding, recall the market rules introduced in Section 6.2. The dispatch algorithm dispatches generators based on their bid prices and the regional trading price is set uniformly for a region based on the average of the six dispatch prices in a trading interval. If a price spike occurs, generators subsequently rebid to floor price in the following dispatch intervals of the same trading period. Since the rebidding is at the floor price, which is lower than the bid price of most dispatched generators in the neighbouring region, these bids are dispatched to replace the generators in the neighbouring region with the result of a change in the direction of inter-regional flow.

Careful inspection of Table 6.2 reveals that a counter-price flow occurred in this trading interval. The inter-regional out-flows shown in the last column indicate that the total out-flow

in that trading interval is 2,028 MW largely due to the change in the direction of inter-regional flow after the spike. In Figure 6.7, it shows that in this trading interval, there is a large scale rebidding to the floor price after the spike. In response to the large amount of low price bids becoming available, the dispatch algorithm then dispatches these bids in Queensland to replace New South Wales generators, thus resulting in a significant increase in dispatched capacity and a change in the direction of inter-regional flow (shown in Table 6.2 as large increases in dispatched quantity in column four, but stable load condition in column three with significant increases of Queensland out-flow in column five). Note that the six dispatch prices shown in Table 6.2 lead to a settlement price of around A$2,223/MWh in Queensland, whereas the trading price in New South Wales is around A$56/MWh at the same period. It follows that this regional outflow from Queensland is a counter-price flow.

6.6 Policy implications

The strategic bidding and rebidding behaviour by base generators in the NEM is evidence of market power that derives primarily from the physical characteristics of the electricity market, namely the difference between the ramp-up and synchronisation rates of generators. This market power can only be exercised because of a failure of regulatory policy. Specifically, a competitive market has been fostered but the rules governing the bidding, dispatch and settlement of electricity have not suitably changed. This has led to a unique situation in which strategic behaviour on the part of the base generators can be very profitable.

The analysis of the preceding two sections has demonstrated that strategic bidding (bid-splitting) and rebidding has undesirable consequences for the electricity market. Strategic bidding can lead to extreme spikes in the dispatch price thus inflating the half-hourly settlement price, which therefore does not reflect the fundamental cost of electricity supply and also makes rebidding after the price spike a profitable enterprise. Rebidding not only allows the base generators to reap the benefit of strategic bidding, but also results in substantial changes in the supply conditions of the regional market for a short period of time and induces counter-price flows. Counter-price flow is of particular concern for the market regulator

(Australian Energy Market Commission, 2013), as it could impact on the hedging effect of inter-regional settlement residues auction, which AEMO introduced as an important risk management tool for market participants (Anderson et al., 2007; Australian Energy Market Operator, 2014).

Despite these undesirable consequences, strategic bidding and rebidding remains possible under the current market rules of the NEM. In fact, as a result of the introduction of the renewable energy target (Kent and Mercer, 2006), base load generators in the NEM face a decreasing market share in the next decade, and may thus be motivated to behave strategically in order to maintain their profitability. Moreover, the consecutive increases in the cap price indicated in Figure 6.1 further encourages the bidding behaviour by making it more profitable. It is clear that raising the cap price is among the series of market-oriented policy changes of the NEM, which intended to create a fully competitive market. However, in this case, the promotion of a competitive electricity market without appropriate adjustments to other regulatory policies has resulted in unintended but damaging side effects. So what are the policy measures that could be considered in order to alleviate the problem while not compromising competition?

1. **Dispatch Algorithm**:

   The current dispatch algorithm of the NEM dispatches generators based strictly on lowest bid price first without considering the time a marginal generator may require to start generation (become synchronised). A more sophisticated algorithm is preferred if it could dispatch the marginal generator in advance based on the information available to the effect that the in-use generators are nearly depleted, at least in terms of the current bid stack. In this way, strategic bidding and rebidding that derives from the short-lived extreme spikes caused by the synchronisation rate of marginal generators would not be possible. Moreover, counter-price flows represent a mismatch between the measurement of cost used by dispatch algorithm (bid price) and the one used by market participants (trading price). Thus, a more appropriate measure of cost is preferable for the dispatch algorithm in terms of preventing counter-price flows.

2. **Regional Prices**:

   As discussed in Section 6.4, the success of strategic bid splitting of large scale base generators depends crucially on the synchronisation rates of marginal generators and limitations of the local transmission network. It may be argued that the fast response of base generators in frequently constrained areas reflects their competitiveness relative to slower response local marginal generators. However, the current rules require that all generators in the region receive the same price so there is no incentive for investment in either intra-regional transmission networks or local generation capacity to mitigate these outcomes. Consequently, allowing for different prices within regions would alleviate this problem and encourage investment.

3. **Price Floor**:

   Changing the dispatch algorithm and pricing setting rules can eradicate the issue of strategic bidding and rebidding, however, they require fundamental changes to the current market structure and introduce significant complications to the operation of the market. A less costly alternative would be lowering the market floor price. Lowering the floor price allows the generators to fully compete with each other in the dispatch intervals after price spikes, and drive down the trading price further. In this way, strategic bidding and rebidding are discouraged due to the lower price incentive. However an undesirable effect is that it also leads to a more radical change in market conditions in a short period and possibly more severe counter-price flows.

4. **Bidding Process**:

   Another more flexible solution would be to change the rebidding rules to allow rebidding only up to 30 minutes before dispatch. In terms of this change, the profitability of strategic bidding would decrease due to not being able to be dispatched in full capacity in the intervals following an abnormally high dispatch price. This change also has the desirable effect of increasing the opportunity cost of strategic bidding by forcing the capacity bid at extremely high price levels to be in place for longer periods. More importantly, changing the bidding rules completely eliminates the possibility

of rebidding after spikes in the same half-hour trading interval and thus decreases the possibility of having large-scale counter-price flows.

Similarly, an alternative policy measure would be to put a longer time limit on the frequency of rebidding allowed, that is to allow generators to change their bid only after a certain amount of time such as 60 or 120 minutes after the last bid change was made. This change also increases the opportunity cost of strategic bidding and rebidding and therefore discourages speculative behaviour by generators.

From the options discussed earlier, a change of the rebidding rule would perhaps be the least disruptive policy change. It should have relatively little impact on generators in the NEM, since they would still have more flexibility in scheduling generation compared with the generators in other deregulated electricity markets such as PJM in the U.S. (more than 24 hours ahead) and Nord Pool spot in Europe (12 hours ahead).

6.7 Conclusion

In pursuing a lower cost electricity supply, most of the major electricity markets in the world have been deregulated by introducing market orientated competition. However, due to the importance of electricity supply to a nation's economy, the security of the supply requires adequate rules to be imposed for regulating the behaviour of market participants. A balance between the security of electricity supply and the free competition allowed then becomes important to regulators when forming the policy. Moreover, the unique physical characteristics of electricity markets, such as non-storability of electricity, balancing equilibria over many locations, transmission constraints and synchronisation rates further contribute to the difficulty of achieving optimal regulation in a deregulated electricity market.

In this chapter, it is shown that at the five-minute dispatch frequency, there are interesting irregularities in the movement of price, inter-regional flow, load and dispatched capacity, which are related to the physical characteristics of electricity markets. More importantly, when a large and detailed dataset at the same five-minute frequency containing the bid information from every generator in the NEM is examined, evidence of strategic bidding

and rebidding is found. In combination with the characteristics of electricity markets, the strategic bidding and rebidding enable large scale base generators to exploit the market under the current policy in the NEM.

Among the findings in this chapter, perhaps the most important one is the rebidding behaviour of base generators after observing a spike, for which, only negative impacts could arise on the operation of the market. In particular, the rebidding generally leads to substantial changes in market conditions for a short period of time until the trading interval with a spike is over. As a result, this contributes directly to the occurrence of counter-price flows. To address these issues, a suitable way would be to change the rebidding rule to at least 30-minute prior to actual dispatch or putting a time limit on the frequency of rebidding allowed.

Overall, the findings show that the price of electricity in a deregulated market is related to a range of issues that are unique to electricity markets, such as transmission constraints, synchronisation rates, strategic bid splitting and rebidding. This interaction between the price of electricity, bidding behaviour of generators and the regulator environment is expected to continue with the regulation process. One interesting question that may be considered in the future is the impact of a new regulatory policy, the renewable energy target. In particular, how the bidding behaviour of base generators changes when the market supply is made up in a significant portion by renewable energy sources such as solar and wind, that are highly dependent on weather conditions.

Chapter 7

Concluding remarks

Electricity markets are one of the most important type of commodity markets that provide fundamental support to the economic growth of a nation. And their well-functioning is therefore of central concern to both policy makers and all the participants in the market. In the pursuit of increased efficiency, most electricity markets around the world have moved towards deregulation. However, the fundamental restructuring associated in the deregulation process not only raises the importance of some issues that already exist in regulated markets but also creates new challenges to both the market regulator and all other market participants. Consequently, the success of the deregulation depends crucially on the understanding of all the issues involved. Among these, load forecasting, price movements and strategic bidding behaviour of generators, which originate from both the demand and supply side of the market, are of great importance.

This thesis started with a demand side problem, load forecasting. More specifically, a multiple equation time-series model is proposed in Chapter 3 for short-term one day ahead load forecasting. It is shown that the proposed model, which is estimated by repeated application of ordinary least squares, has the potential to match or even outperform more complex nonlinear and nonparametric forecasting models. The key ingredient of the success of this simple model is the effective use of lagged information by allowing for interaction between seasonal patterns and intra-day dependencies. Although the model is built using data for the Queensland region of Australia, the methods are completely generic and applicable

to any load forecasting problem. The model's forecasting ability is assessed by means of the mean absolute percentage error (MAPE). For day ahead forecasts, the MAPE returned by the model over a period of 11 years is an impressive 1.36%. The forecast accuracy of the model is compared with a number of benchmarks including three popular alternatives and one industrial standard reported by the Australia Energy Market Operator (AEMO). The performance of the proposed model is superior to all benchmarks and outperforms the AEMO forecasts by about a third in terms of the MAPE criterion.

In order to provide a more complete picture about the decision risk involved when relying on the load forecasts, Chapter 4 presents two models for forecasting the quantiles of load. The two proposed models for forecasting quantiles of load are formulated in the Bayesian framework. In the first model, a mixture representation of the asymmetric Laplace distributed random variable is used with a computationally efficient Gibbs sampler. The advantage of this approach is not only computational but also robust for the reason that all parameters in the conditional distributions have closed forms. As a consequence, satisfactory results can be obtained even with a large number of parameters. This aspect of the method is particularly useful in terms of forecasting the quantiles of load due to the ability to account for complex seasonal patterns and a large number of important covariates. In the second approach, a non-parametric density is used to replace the corresponding density of ALD and thus relax the stringent distributional assumption of ALD. The results show that the performance is not preferable at least in terms of the two criterion (coverage ratio and pin-ball loss) considered. Overall, the proposed model with ALD is both computationally efficient and very competitive when comparing with the second approach and the AEMO forecasts. However, the conclusion drawn is that the real preference of one model over another should be based on the problem at hand. The approach of using ALD penalises large residuals more severely in one tail than the other when the quantile of interest differs from 0.5. Whereas the non-parametric approach does not have this characteristic. Therefore, the decision on the choice of the models should be made based on whether over-prediction is preferable to under-prediction. In the context of load forecasting, the proposed model with ALD is preferable.

Moving attention to the general equilibrium of a electricity market, the price of electricity, as the main indicator of the overall market equilibrium, is studied in detail in Chapter 5 with the focus on the effect from transmission constraints. The results show that transmission constraints as a unique physical characteristic of electricity markets, contribute significantly to price variability and are a necessary condition for the occurrence of extreme prices. Furthermore, it is found that in the presence of extremely high prices brought about by the presence of constraints, robust estimation techniques are necessary to guard against incorrect inference. It follows that ignoring transmission constraints when modelling electricity prices can severely distort measures of price risk.

Following the study on prices in the previous chapter, Chapter 6 investigates the impact of the extreme price events and other market conditions on the supply side of the market. Specifically, the last part of this thesis, Chapter 6, is focused on the strategic bidding behaviour of generators with the question addressed from a regulatory point of view. In a regulatory setting, the success of deregulation in electricity markets depends crucially on the efficacy of market rules in promoting and maintaining free competition. There are, however, important physical characteristics in electricity markets, such as the constraints imposed by the physical transmission infrastructure, that have the potential to undermine competition and provide opportunities for market participants to exploit temporary positions of market power. Focussing on the regional market of Queensland, Australia is analysed. It is found that strategic behaviour by generators exists and is closely related to the occurrence of the kind of extreme price events that characterise many deregulated electricity markets. In addition, rebidding behaviour by base-load generators immediately after extreme price events is shown to have negative impacts on the operation of the market. The key message is that promotion of competition in electricity markets can have undesirable consequences unless the regulatory policy is carefully designed to counter strategic behaviour by market participants.

In summary, this thesis has provided a detailed investigation into three major issues in deregulated electricity markets, namely, load forecasting, price movements and strategic behaviour by generators. However, as the deregulation process extends into the future, a

continuous effort is required for deepening the understanding of the operation of deregulated electricity markets. In the study on point forecasting of load in Chapter 3, the use of a simple measure of accuracy, namely, the mean absolute percentage error enables straightforward model development and comparison. However, the simple computation of this error metric does not really encapsulate the economic advantage to market participants of providing accurate load forecasts. The challenge for future work is therefore to devise a metric that is capable of measuring economic gains to more accurate load forecasting. Whereas the study of quantile load forecasting in Chapter 4 only represents a starting point in this area, a more flexible quantile regression framework needs to be devised in future work in order to accurately represent individual attitudes toward risk. Moreover, from the studies in Chapters 5 and 6, the behaviour of electricity price in a deregulated market is shown to be related to a range of issues that are unique to electricity markets, such as transmission constraints, synchronisation rates, strategic bid splitting and rebidding. Also, it is only at the five-minute frequency that these issues became clear. Traditionally, attention has been mainly focused on half-hour trading prices, using traditional covariates such as load, lagged price and reserve margin. Consequently, a fertile area of future research in modelling electricity prices would be to turn to the five-minute frequency where new supply side data becomes important.

References

Aderounmu, A. A. and Wolff, R. (2014). Modeling dependence of price spikes in Australian electricity markets. *Energy Risk*, 11:60–65.

Albanese, C., Lo, H., and Tompaidis, S. (2012). A numerical algorithm for pricing electricity derivatives for jump-diffusion processes based on continuous time lattices . *European Journal of Operational Research*, 222:361–368.

Amaral, L. F., Souza, R. C., and Stevenson, M. (2008). A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting. *International Journal of Forecasting*, 24:603–615.

Anderson, C. L. and Davison, M. (2008). A hybrid system-econometric model for electricity spot prices: considering spike sensitivity to forced outage distributions. *IEEE Transactions on Power Systems*, 23:927–937.

Anderson, E. J., Hu, X., and Winchester, D. (2007). Forward contracts in electricity markets: The Australian experience. *Energy Policy*, 35:3089–3103.

Areekul, P., Senju, T., Toyama, H., Chakraborty, S., Yona, A., Urasaki, N., Mandal, P., and Saber, A. Y. (2010). A new method for next-day price forecasting for PJM electricity market. *International Journal of Emerging Electric Power Systems*, 11.

Australian Energy Market Commission (2013). The National Electricity Market: A case study in microeconomic reform. (Online: http://www.aemc.gov.au/About-Us/Resources/Corporate-publications/The-National-Electricity-Market-A-case-study-in-mi. accessed January 2016).

Australian Energy Market Commission (2015). National Electricity Rules Version 71. (Online: http://www.aemc.gov.au/Energy-Rules/National-electricity-rules/Current-Rules. accessed April 2015).

Australian Energy Market Operator (2014). Guide to the Settlements Residue Auction. (Online: http://www.aemo.com.au/Electricity/Market-Operations/Settlement-Residue-Auction/Guide. accessed April 2015).

Australian Energy Regulator (2015). State of the Energy Market 2015. Technical report, Australian Energy Regulator.

Barlow, M. T. (2002). A DIFFUSION MODEL FOR ELECTRICITY PRICES. *Mathematical Finance*, 12:287–298.

Becker, R., Hurn, S., and Pavlov, V. (2007). Modelling spikes in electricity prices. *Economic Record*, 83:371–382.

Benth, F. E., Kallsen, J., and Meyer-Brandis, T. (2007). A Non-Gaussian Ornstein-Uhlenbeck Process for Electricity Spot Price Modeling and Derivatives Pricing. *Applied Mathematical Finance*, 14:153–169.

Bhar, R., Colwell, D. B., and Xiao, Y. (2013). A jump diffusion model for spot electricity prices and market price of risk. *Physica A: Statistical Mechanics and its Applications*, 392:3213–3222.

Biggar, D. (2011). The Theory and Practice of the Exercise of Market Power in the Australian NEM. Technical report, Australian Energy Market Commission.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson Distributions Via Pólya Urn Schemes. *The Annals of Statistics*, 1:353–355.

Bosco, B., Parisio, L., and Pelagatti, M. (2012). Strategic bidding in vertically integrated power markets with an application to the Italian electricity auctions. *Energy Economics*, 34:2046–2057.

Bowden, N. and Payne, J. E. (2008). Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models. *Energy Economics*, 30:3186–3197.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Briggs, R. J. and Kleit, A. (2013). Resource adequacy reliability and the impacts of capacity subsidies in competitive electricity markets. *Energy Economics*, 40:297–305.

Bunn, D. W. and Martoccia, M. (2005). Unilateral and collusive market power in the electricity pool of England and Wales. *Energy Economics*, 27:305–315.

Burnett, J. W. and Zhao, X. (2015). "Spatially Explicit Prediction of Wholesale Electricity Prices". *International Regional Science Review*, pages 1–42.

Bustos-Salvagno, J. (2015). Bidding behavior in the Chilean electricity market. *Energy Economics*, 51:288–299.

Byström, H. N. E. (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*, 14:41–55.

Cancelo, J. R., Espasa, A., and Grafe, R. (2008). Forecasting the electricity load from one day to one week ahead for the Spanish system operator. *International Journal of Forecasting*, 24:588–602.

Cartea, A. and Figueroa, M. G. (2005). Pricing in Electricity Markets: A Mean Reverting Jump Diffusion Model with Seasonality. *Applied Mathematical Finance*, 12:313–335.

Cartea, Á., Figueroa, M. G., and Geman, H. (2009). Modelling electricity prices with forward looking capacity constraints. *Applied Mathematical Finance*, 16:103–122.

Cepeda, M. and Finon, D. (2011). Generation capacity adequacy in interdependent electricity markets. *Energy Policy*, 39:3128–3143.

Chaâbane, N. (2014). A novel auto-regressive fractionally integrated moving average–least-squares support vector machine model for electricity spot prices prediction. *Journal of Applied Statistics*, 41:635–651.

Chan, K. F. and Gray, P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*, 22:283–300.

Chan, V., Lahiri, S. N., and Meeker, W. Q. (2004). Block Bootstrap Estimation of the Distribution of Cumulative Outdoor Degradation. *Technometrics*, 46:215–224.

Chernozhukov, V. and Fernández-Val, I. (2005). Subsampling inference on quantile regression processes. *Sankhyā: The Indian Journal of Statistics*, 67:253–276.

Christensen, T., Hurn, S., and Lindsay, K. (2009). It never rains but it pours: modelling the persistence of spikes in electricity prices. *The Energy Journal*, 30:25–48.

Christensen, T. M., Hurn, A. S., and Lindsay, K. A. (2012). Forecasting spikes in electricity prices. *International Journal of Forecasting*, 28:400–411.

Christoffersen, P. and Pelletier, D. (2004). Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics*, 2:84–108.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39:841–862.

Clements, A., Fuller, J., and Hurn, S. (2013). Semi-parametric Forecasting of Spikes in Electricity Prices. *Economic Record*, 89:508–521.

Clements, M. P., Galvão, A. B., and Kim, J. H. (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*, 15:729–750.

Contreras, J., Espinola, R., Nogales, F., and Conejo, A. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, 18:1014–1020.

Cottet, R. and Smith, M. (2003). Bayesian Modeling and Forecasting of Intraday Electricity Load. *Journal of the American Statistical Association*, 98:839–849.

Cruz, A., Muñoz, A., Zamora, J. L., and Espínola, R. (2011). The effect of wind generation and weekday on Spanish electricity spot price forecasting . *Electric Power Systems Research*, 81:1924–1935.

Cuaresma, J. C., Hlouskova, J., Kossmeier, S., and Obersteiner, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models . *Applied Energy*, 77:87–106.

Darbellay, G. A. and Slama, M. (2000). Forecasting the short-term demand for electricity: Do neural networks stand a better chance? *International Journal of Forecasting*, 16:71–83.

de Jong, C. and Huisman, R. (2003). Option pricing for power prices with spikes. *Energy Power Risk Management*, 7:12–16.

De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106:1513–1527.

Ding, F. and Fuller, J. D. (2005). Nodal, uniform, or zonal pricing: distribution of economic surplus. *IEEE Transactions on Power Systems*, 20:875–882.

Douglas, S. M. and Popova, J. N. (2011). Econometric Estimation of Spatial Patterns in Electricity Prices. *The Energy Journal*, 32:81–105.

Eichler, M. and Tuerk, D. (2013). Fitting semiparametric Markov regime-switching models to electricity spot prices. *Energy Economics*, 36:614–624.

Engle, R. F., Chowdhury, M., and Rice, J. (1992). Modelling peak electricity demand. *Journal of Forecasting*, 11:241–251.

Engle, R. F., Granger, C. W. J., and Hallman, J. J. (1989). Merging short-and long-run forecasts: An application of seasonal cointegration to monthly electricity sales forecasting. *Journal of Econometrics*, 40:45–62.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics*, 22:367–381.

Escobar, M. D. and West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, 90:577–588.

Escribano, A., Ignacio Peña, J., and Villaplana, P. (2002). Modelling electricity prices: International evidence. Working paper, Universidad Carlos III de Madrid.

Espinoza, M., Joye, C., Belmans, R., and De Moor, B. (2005). Short-Term Load Forecasting, Profile Identification, and Customer Segmentation: A Methodology Based on Periodic Time Series. *IEEE Transactions on Power Systems*, 20:1622–1630.

Fan, S. and Hyndman, R. J. (2012). Short-Term Load Forecasting Based on a Semi-Parametric Additive Model. *IEEE Transactions on Power Systems*, 27:134–141.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230.

Fitzenberger, B. (1998). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics*, 82:235–287.

Garcia, R. C., Contreras, J., Van Akkeren, M., and Garcia, J. B. C. (2005). A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, 20:867–874.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian data analysis*. Chapman & Hall, London, 3rd edition.

Gerlach, R. H., Chen, C. W. S., and Chan, N. Y. C. (2011). Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets. *Journal of Business & Economic Statistics*, 29:481–492.

Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., and Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, 191:207–222.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242.

Hagan, M. T. and Behr, S. M. (1987). The Time Series Approach to Short Term Load Forecasting. *IEEE Transactions on Power Systems*, 2:785–791.

Harvey, A. and Koopman, S. J. (1993). Forecasting Hourly Electricity Demand Using Time-Varying Splines. *Journal of the American Statistical Association*, 88:1228–1236.

Higgs, H. and Worthington, A. (2008). Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The Australian wholesale spot electricity market. *Energy Economics*, 30:3172–3185.

Hippert, H. S., Pedreira, C. E., and Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16:44–55.

Hu, X., Grozev, G., and Batten, D. (2005). Empirical observations of bidding patterns in Australia's National Electricity Market. *Energy Policy*, 33:2075–2086.

Huisman, R. and Mahieu, R. (2003). Regime jumps in electricity prices. *Energy Economics*, 25:425–434.

Huurman, C., Ravazzolo, F., and Zhou, C. (2012). The power of weather. *Computational Statistics & Data Analysis*, 56:3793–3807.

Hyndman, R. J. and Fan, S. (2010). Density Forecasting for Long-Term Peak Electricity Demand. *Power Systems, IEEE Transactions on*, 25:1142–1153.

Ishwaran, H. and James, L. F. (2001). Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association*, 96:161–173.

Janczura, J., Trück, S., Weron, R., and Wolff, R. C. (2013). Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics*, 38:96–110.

Jorion, P. (2007). *Value at risk: the new benchmark for managing financial risk*. McGraw-Hill, New York.

Kanamura, T. and Ohashi, K. (2008). On transition probabilities of regime switching in electricity prices. *Energy Economics*, 30:1158–1172.

Karakatsani, N. V. and Bunn, D. W. (2008). Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24:764–785.

Karthikeyan, S. P., Raglend, I. J., and Kothari, D. (2013). A review on market power in deregulated electricity market. *International Journal of Electrical Power & Energy Systems*, 48:139–147.

Kent, A. and Mercer, D. (2006). Australia's mandatory renewable energy target (MRET): an assessment. *Energy Policy*, 34:1046–1062.

Kim, M. S. (2013). Modeling special-day effects for forecasting intraday electricity demand. *European Journal of Operational Research*, 230:170–180.

Knittel, C. R. and Roberts, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, 27:791–817.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.

Koenker, R. and Bassett Jr., G. (1978). Regression Quantiles. *Econometrica*, 46:33–50.

Koenker, R. and Xiao, Z. (2006). Quantile Autoregression. *Journal of the American Statistical Association*, 101:980–1006.

Koschker, S. and Möst, D. (2015). Perfect competition vs. strategic behaviour models to derive electricity prices and the influence of renewables on market power. *OR Spectrum*, pages 1–26.

Kottas, A. and Krnjajić, M. (2009). Bayesian Semiparametric Modelling in Quantile Regression. *Scandinavian Journal of Statistics*, 36:297–319.

Kotz, S., Kozubowski, K., and Podgorski, K. (2001). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Birkhäuser, Boston.

Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81:1565–1578.

Liu, B., Nowotarski, J., Hong, T., and Weron, R. (2015). Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts. *IEEE Transactions on Smart Grid*, pages 1–8.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In Lepage, R., Billard, L., and Nelson, R., editors, *Exploring the limits of bootstrap*, pages 225–248. Wiley, New York.

López Cabrera, B. and Schulz, F. (2014). Forecasting generalized quantiles of electricity demand: A functional data approach. Technical report, SFB 649 Discussion Paper.

Lucia, J. J. and Schwartz, E. S. (2002). Electricity prices and power derivatives: Evidence from the Nordic power exchange. *Review of Derivatives Research*, 5:5–50.

Lízal, L. M. and Tashpulatov, S. N. (2014). Do producers apply a capacity cutting strategy to increase prices? The case of the England and Wales electricity market. *Energy Economics*, 43:114–124.

Maenhoudt, M. and Deconinck, G. (2014). Strategic offering to maximize day-ahead profit by hedging against an infeasible market clearing result. *IEEE Transactions on Power Systems*, 29:854–862.

Mandal, P., Senjyu, T., and Funabashi, T. (2006). Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy Conversion and Management*, 47:2128–2142.

McSharry, P. E., Bouwman, S., and Bloemhof, G. (2005). Probabilistic forecasts of the magnitude and timing of peak electricity demand. *IEEE Transactions on Power Systems*, 20:1166–1172.

Misiorek, A., Trueck, S., and Weron, R. (2006). Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. *Studies in Nonlinear Dynamics & Econometrics*, 10.

Mount, T. D. and Ju, J. (2014). An econometric framework for evaluating the efficiency of a market for transmission congestion contracts. *Energy Economics*, 46:176–185.

Mount, T. D., Ning, Y., and Cai, X. (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, 28:62–80.

Nappu, M. B., Bansal, R. C., and Saha, T. K. (2013). Market power implication on congested power system: A case study of financial withheld strategy. *International Journal of Electrical Power & Energy Systems*, 47:408–415.

Outhred, H. (2000). The competitive market for electricity in Australia: why it works so well. In *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*, pages 1–8. IEEE.

Pardo, A., Meneu, V., and Valor, E. (2002). Temperature and seasonality influences on Spanish electricity load. *Energy Economics*, 24:55–70.

Park, D. C., El-Sharkawi, M. A., Marks, R. J., Atlas, L. E., and Damborg, M. J. (1991). Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, 6:442–449.

Patton, A., Politis, D. N., and White, H. (2009). Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews*, 28:372–375.

Peirson, J. and Henley, A. (1994). Electricity load and temperature: Issues in dynamic specification. *Energy Economics*, 16:235–243.

Pindoriya, N. M., Singh, S. N., and Singh, S. K. (2008). An adaptive wavelet neural network-based energy price forecasting in electricity markets. *IEEE Transactions on Power Systems*, 23:1423–1432.

Politis, D. N. (2001). Resampling time series with seasonal components. In *Frontiers in data mining and bioinformatics: Proceedings of the 33rd symposium on the interface of computing science and statistics*, pages 619–621, Orange County, California.

Powell, J. L. (1991). Estimation of Monotonic Regression Models Under Quantile Restrictions. In Barnett, W. A., Powell, J., and Tauchen, G. E., editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, pages 357–384. Cambridge University Press, Cambridge.

Ramanathan, R., Engle, R., Granger, C. W. J., Vahid-Araghi, F., and Brace, C. (1997). Shorte-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, 13:161–174.

Robinson, T. A. (2000). Electricity pool prices: a case study in nonlinear time-series modelling. *Applied Economics*, 32:527–532.

Sansom, D. C., Downs, T., and Saha, T. K. (2002). Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. *Journal of Electrical and Electronics Engineering*, 22:227–233.

Schwenen, S. (2014). Market design and supply security in imperfect power markets. *Energy Economics*, 43:256–263.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650.

Short, C. and Swan, A. (2002). *Competition in the Australian national electricity market*. ABARE.

Simshauser, P. (2010a). Resource adequacy, capital adequacy and investment uncertainty in the Australian power market. *The Electricity Journal*, 23:67–84.

Simshauser, P. (2010b). Vertical integration, credit ratings and retail price settings in energy-only markets: Navigating the Resource Adequacy problem. *Energy Policy*, 38:7427–7441.

Soares, L. J. and Medeiros, M. C. (2008). Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. *International Journal of Forecasting*, 24:630–644.

Spliid, H. (1983). A Fast Estimation Method for the Vector Autoregressive Moving Average Model With Exogenous Variables. *Journal of the American Statistical Association*, 78:843–849.

Srinivasan, D., Chang, C. S., and Liew, A. C. (1995). Demand forecasting using fuzzy neural computation, with special emphasis on weekend and public holiday forecasting. *IEEE Transactions on Power Systems*, 10:1897–1903.

Tamaschke, R., Docwra, G., and Stillman, R. (2005). Measuring market power in electricity generation: A long-term perspective using a programming model. *Energy Economics*, 27:317–335.

Taylor, J. W. (2007). Forecasting daily supermarket sales using exponentially weighted quantile regression . *European Journal of Operational Research*, 178:154–167.

Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204:139–152.

Taylor, J. W. (2012). Short-Term Load Forecasting With Exponentially Weighted Methods. *IEEE Transactions on Power Systems*, 27:458–464.

Taylor, J. W. and Buizza, R. (2002). Neural network load forecasting with weather ensemble predictions. *IEEE Transactions on Power Systems*, 17:626–632.

Taylor, J. W. and McSharry, P. E. (2007). Short-Term Load Forecasting Methods: An Evaluation Based on European Data. *IEEE Transactions on Power Systems*, 22:2213–2219.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30:1030–1081.

Weron, R., Bierbrauer, M., and Trück, S. (2004). Modeling electricity prices: jump diffusion and regime switching. *Physica A: Statistical Mechanics and its Applications*, 336:39–48.

Weron, R. and Misiorek, A. (2008). Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, 24:744–763.

Wolak, F. A. (2000). An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market. *International Economic Journal*, 14:1–39.

Yamin, H. Y., Shahidehpour, S. M., and Li, Z. (2004). Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. *International journal of electrical power & energy systems*, 26:571–581.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54:437–447.

Zareipour, H., Cañizares, C. A., Bhattacharya, K., and Thomson, J. (2006). Application of public-domain market information to forecast Ontario's wholesale electricity prices. *IEEE Transactions on Power Systems*, 21:1707–1717.

Zareipour, H., Janjani, A., Leung, H., Motamedi, A., and Schellenberg, A. (2011). Classification of future electricity market prices. *IEEE Transactions on Power Systems*, 26:165–173.

Zhang, B., Johari, R., and Rajagopal, R. (2015). Competition and coalition formation of renewable power producers. *IEEE Transactions on Power Systems*, 30:1624–1632.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14:35–62.

Zhang, N. (2009). Generators' bidding behavior in the NYISO day-ahead wholesale electricity market. *Energy Economics*, 31:897–913.

Zhao, J. H., Dong, Z. Y., Xu, Z., and Wong, K. P. (2008). A statistical approach for interval forecasting of the electricity price. *IEEE Transactions on Power Systems*, 23:267–276.