ELECTRICITY AUTHORITY
TE MANA HIKO

Guide to Market Performance Enquiry: Demand Side Bidding and Forecasting

Market performance enquiry

15 September 2016

Investigation stages

Market Performance enquiries, reviews and investigations

The Electricity Authority (Authority) may carry out an enquiry, review or investigation as a result of monitoring the industry or the market, or at the request of an external party. The Minister of Energy and Resources (Minister) may also ask or direct the Authority to look into an issue.

An enquiry, review or investigation looks at the circumstances giving rise to an out of the ordinary event, including the actions of participants. An enquiry, review or investigation may result in suggestions for Code amendments, market facilitation measures, or in a finding that no further action is needed. In these cases, the Authority usually publishes a report of its findings.

At the same time as it carries out a market performance enquiry, investigation or review, the Authority's compliance team may investigate whether there has been a breach of the Code, Act or Regulations. The two processes may run concurrently, but will not always finish at the same time.

Enquiries, reviews and investigations represent three stages in an escalating process, with increased effort and significance attached to each one.

Market Performance Enquiry (Stage I): At the first stage, the Authority carries out analysis using existing data and resources. The purpose of an enquiry is to better understand circumstances, observed through routine monitoring, that appear to require closer inspection. The Authority will usually announce it is carrying out an enquiry.

If the results of the enquiry show that the circumstances are unlikely to have any implications for the Authority's statutory objective (to promote competition in, reliable supply by, and the efficient operation of, the electricity industry for the long-term benefit of consumers) the Authority is unlikely to take further action. The Authority will publish the results of its enquiry, which may consist of an enquiry paper, or a short note, on the results on the website.

Market Performance Review (Stage II): The Authority will initiate a review if, at the end of a Stage 1 enquiry, it does not have enough information to understand the issue but it appears to be significant for competition in, reliable supply by or efficient operation of the electricity industry for the long-term benefits of consumers. The Authority generally asks relevant service providers and industry participants to supply information. There is typically a period of iterative information-gathering and analysis. The Authority will announce it is undertaking a review, and publish the results.

Market Performance Investigation (Stage III): At this stage, the Authority may exercise its statutory information-gathering powers under section 46 of the Electricity Industry Act to acquire the information it needs to investigate an issue in depth. The Authority will announce early in the process that it is undertaking an investigation and indicate when it expects to complete the work. The Authority will publish reports of Stage III investigations.

Contents

Introduction

1.1 This is a guide to the post-implementation review of the Authority's demand-side bidding and forecasting initiative (DSBF). This guide summarises the methodology used to evaluate DSBF and the results of the evaluation. Further detail on this post-implementation review can be found in the accompanying technical paper.

1.2 The purpose of post-implementation reviews is to assess the Authority's previous decisions and thereby help improve decision-making. The Authority uses cost benefit analysis to help make regulatory decisions, so these analyses form the basis of the post-implementation review.

Background

1.3 Prior to DSBF, all purchasers were required to submit demand bids. These were matched with generator offers and transmission information to create the pre-dispatch schedule (PDS) and the schedule of dispatch prices and quantities (SDPQ). The PDS was a schedule of forecast load and prices and it was produced every two hours for the current day and from 1 pm, every two hours for the following day. SDPQ was a shorter term forecast which was produced every 30 minutes from 4 hours prior to real time. These changes are summarised in Table 1 at the end of this paper.

1.4 SDPQ did not reflect the impact of demand response to price; so when SDPQ signalled a high price it assumed zero demand response. In addition, large consumers were not aware if and by how much their demand response would affect prices. Inefficiency in the market was observed when the participants reduced demand in response to high forecast prices. On occasion, this led to more load being cut than was required.

1.5 In addition to including information related to demand response, the DSBF initiative was expected to deliver improved forecast price accuracy through:

(a) improved load forecasting processes

(b) longer scheduling period and increased scheduling frequency.

Under the DSBF initiative, the Authority determines grid exit points (GXPs) as either "conforming" or "non-conforming"; where conforming GXPs are those whose loads can be forecast accurately by the system operator. Purchasers at non-conforming nodes are required to offer nominated bids indicating the extent to which they will reduce load in response to price. DSBF produces two pre-dispatch price schedules: the non-response schedule (NRS) and the price responsive schedule (PRS), the latter accounting for the demand response due to nominated bids. Having these two distinct schedules—one with demand response to price and one without—helps participants see the combined effect of demand response.

Summary of results

1.6 We do not expect forecasts to be perfect, so we compared different forecasts with each other to see which was better. We found that the PRS is superior to the NRS across a number of measures for both single and multiple forecast horizons. We conclude from the pairwise comparison of the forecasts that both the NRS and the PRS perform better than the a simple time series forecast (ARIMA) forecasts. The difference in forecasting performance is statistically significant.

1.7 We found that the SDPQ performed better than the NRS (these are essentially the same forecasting algorithm however they apply over different time periods). This result could be due to the market responding to the PRS, therefore making the NRS less accurate.

1.8 We looked at extreme values and find that they are more likely in the 2-hour forecasts than in final prices, but less likely in the 30-minute forecasts than in final prices. In other words, the 2-hour schedules over-forecast extreme points and 30 minute schedules under-forecast extreme values.

The cost and benefit analysis

1.9 The cost benefit analysis carried out by the Authority in 2011 identified the following factors as possible benefits of DSBF:

a. enabling increased demand response to price

b. improved system security

c. reduced frequency keeping cost

d. reduced compliance and risk management costs

e. increased dynamic efficiency.

1.10 These benefits were estimated to be $1,337,000 over eight years (mid-point) but are difficult to directly measure. We can assume that compliance costs have reduced simply because consumers at fewer nodes are required to submit nominated bids. However, the remaining benefits are all due to greater coordination between supply and demand. On the supply side, better forecasts should assist unit commitment and fuel purchase decisions. On the demand side, better forecasts should lead to more efficient consumption decisions. If this improved coordination were to flow through to the spot market, the change in prices would be too small to reliably measure.

Approach

1.11 Instead of assessing the expected benefits of DSBF directly, we focused on assessing the accuracy of DSBF forecasts. We did this because the expected benefits of DSBF will only be realised through coordination in the spot market that in turn relies on the forecasts. In a market, the price is the mechanism that coordinates supply and demand. In the spot market, the final price is determined a day after the trading period has occurred, which means decisions about production (unit commitment, fuel choices) and consumption are made before final prices are known. Coordination in the spot market therefore relies on forecast prices. Therefore this post implementation review focuses on the quality of the DSBF forecasts and makes the assumption that if the forecasts have improved, then coordination in the market will improve and the benefits will be realised.

1.12 This paper concerns the Authority's evaluation of the forecasting performance of each of the pre-dispatch schedules. We are particularly interested in relative performance of the NRS against the PRS to see if including demand response has resulted in better forecasts. The forecasting performance of the NRS is also compared with its predecessor in order to determine whether DSBF brought about the improvement in forecasts available to the market. Finally, we build an auto-regressive moving average (ARIMA) model[1] of 5-minute real-time prices to produce a time series forecast. We study

[1] ARIMA is integrated autoregressive and moving average model that explains the movement of forecast errors by relating them to their own past values and to a weighted sum of the current and past random disturbances.

whether an ARIMA forecast produces better forecasts than the DSBF pre-dispatch schedules. ARIMA relates past information to forecast prices in a linear random fashion. Pre-dispatch prices are produced using market information and the market dispatch model. We would expect pre-dispatch prices to perform better than simple ARIMA forecasting.

Data used

1.13 We first investigate the forecasting performance of the individual schedules at Benmore, Otahuhu and Haywards. The 5-minute real-time prices were used to produce ARIMA forecasts which were compared to the PRS for accuracy. Transpower provided data on the SDPQ which pre-dated DSBF and which is compared with the NRS. We removed outliers because this improves the accuracy of the results and reduces the errors of inference by avoiding the distortion of parameter and statistic estimates. We later include the outliers and evaluate the behaviour of extreme values.

Methodology and results

1.14 Time series forecast evaluation analysis is used throughout this review. Our goal is to judge whether the pre-dispatch schedules use available information efficiently in making forecasts and whether one forecasting method is superior to another. In order to do this, we carry out a range of statistical tests – each one tests an attribute of forecasting error, but we need to have a look at a range of tests since there is no one test to check for these properties simultaneously.

1.15 We start with looking at individual schedules and examining the relationship between actual price and forecast using scatter plots. The first few 30-minute forecasts align well with the actual price but differences[2] appear and grow with lead time[3].

Individual forecasts are imperfect

1.16 The first step in our analysis is to test whether the forecast error has a median of zero. This is an intuitively obvious test to see if the factors that are not captured by the forecast are random. If so, then the median forecast error should be zero. We use statistical tests to test this and find evidence that there is non-random information that is excluded from the forecast.

1.17 We look further into the relationship between forecast and actual prices, and at the individual characteristics of each forecast by examining various distributions. We use both quantitative and graphical measures.

1.18 The correlation coefficient ranges between 1.0, where there is an exact linear relationship between forecasts and final price, and zero, where there is no relationship. The correlation coefficient describes the quality or relative accuracy of the forecasts in the sample. It appears that the correlation between final price and forecast overall reduces as lead time lengthens; indicating that, as a trading period approaches, the forecast improves. This is intuitive as it indicates that uncertainty is resolved as time passes.

1.19 We next examine the degree to which forecast are too high or too low using a statistical method based on regressing the final prices on the forecasts. This involves testing whether the value of the constant equals zero and the value of the regression coefficient

[2]Technically this is called bias which means either the forecasts are typically too high or too low.

[3] Lead time is the time between the current and forecasted period.

equals 1.0. We conclude from this analysis that all the forecasts exhibit errors that seems systematic. The PRS and NRS schedules under-forecast low prices and over-forecast high prices. This means that forecasts are typically too low for low prices and too high for high prices.

1.20 We also find that the forecasts are not making the best use of available information. This is formally tested by using a set of statistical methodology in the main paper. We find that the forecasts are not optimal and there is a room for improvement in producing predispatch schedules.

1.21 Multi-horizon optimality is also examined. We explicitly take into account multi-horizon relationship to minimise the risk of mistakenly concluding optimality based on single forecast evaluation. We find evidence for non-optimality in all the forecasts. We could use information more efficiently in making forecasts.

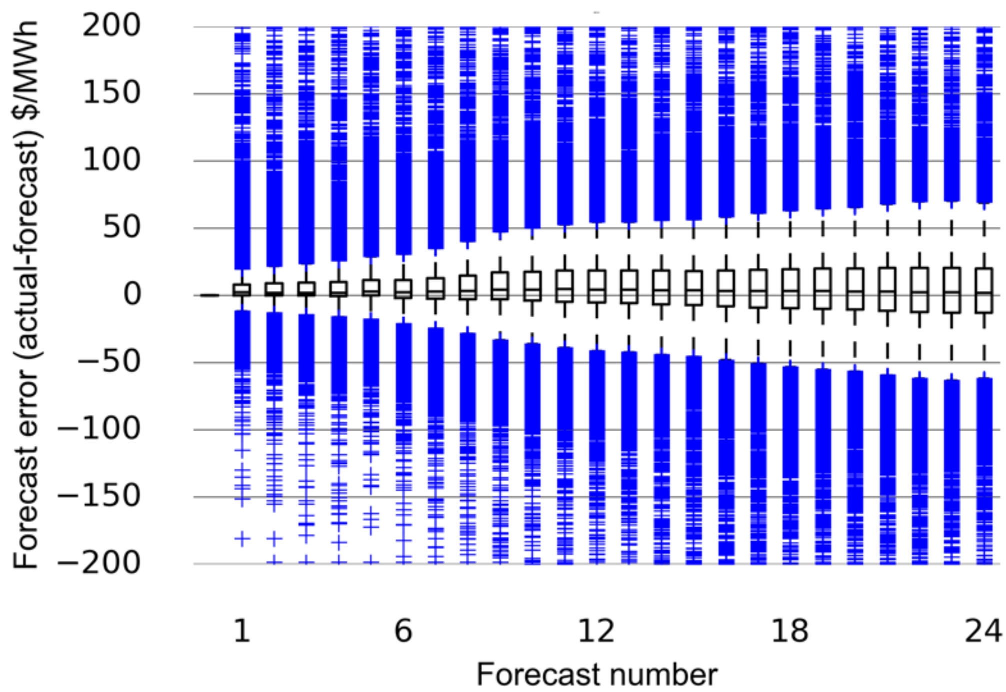
Multiple forecasts comparison: the PRS performs better overall

1.22 Measures of forecast accuracy are convenient tools to arrive at rankings of forecast ability for multiple forecasts. Because all the forecasts are on the same scale we can safely use RMSE and MAPE which are scale dependent. RMSE is the square root of mean of forecast error squared and MAPE is the mean of the absolute values of percentage error. RMSE measures the average magnitude of the error while MAPE measures the same thing in percentage terms.

1.23 According to these measures of forecast accuracy, forecasts with shorter lead times rank better than those with longer lead times. This is expected since the accuracy should improve closer to the real time as uncertainty is resolved. This could also be partly due to how Transpower responds to forecast prices. The generators increase offered quantities in response to high prices and Transpower's control centre operators push the load forecast higher to ensure 30 minute peak load can be covered.

1.24 The PRS performs better than the NRS and Benmore schedules are better than Otahuhu and Haywards, except the 2-hour forecasts of Haywards. The poorest performers overall are the Otahuhu schedules. The PRS' greater accuracy may be due to traders following their nominated bids.

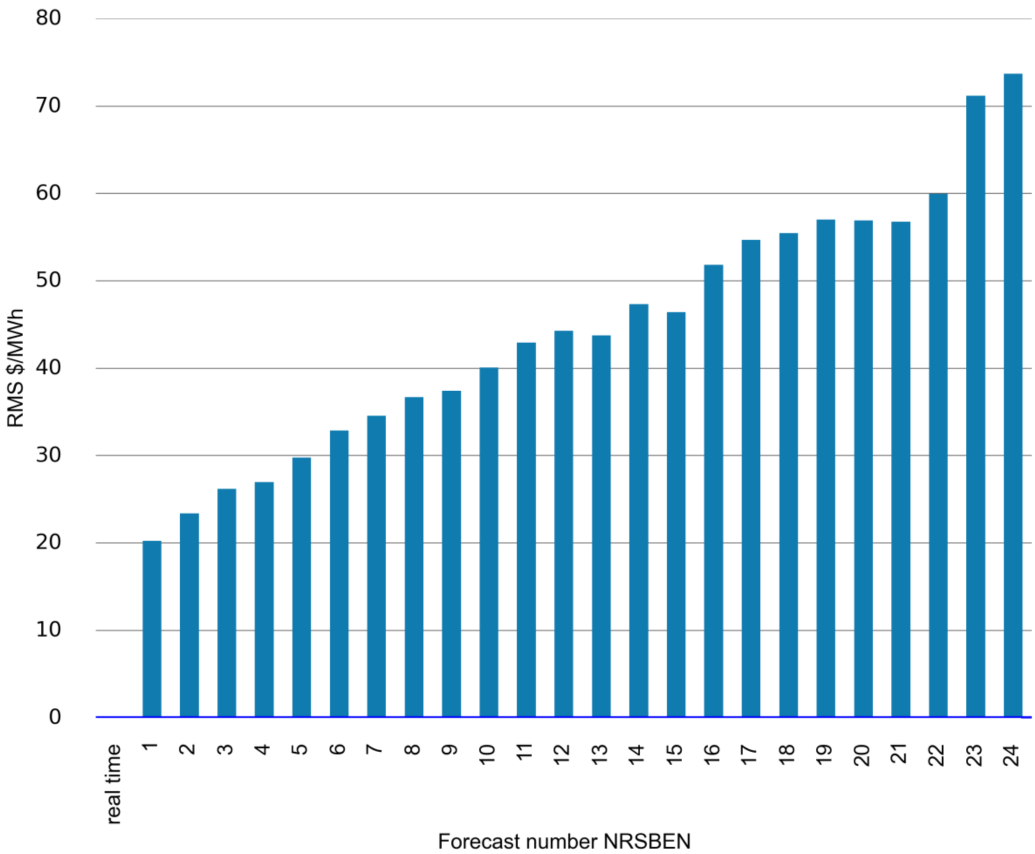
Figure 1 Boxplot of forecast error for the NRS Benmore



1.25 Figure 1 shows Box-Whisker plots for forecast error in an attempt to illustrate the shape of the error measure distributions. These show central tendency, variability[4] and symmetry of distribution of forecast errors. The x-axis lists forecast horizons with 1 to 8 being 30-minute schedules which run up to 4 hours from real time. Forecasts 9 to 24 are 2-hour forecasts which run from 4 to 36 hours out from real time. The lines shaded blue are outliers and the length of the boxes are interquartile range; these show the variability of the forecasts. The top line of the box is upper quartile and the bottom line of the box is lower quartile. The line inside the box is median, showing the central value of the distribution. The distribution is positively skewed with many extreme values—the y-axis is truncated to give an accurate picture of the convergence. The central tendency is similar across all the forecast schedules and the chart shows that variability increases with lead time. We can clearly see the convergence of the distribution towards forecasts with shorter lead time.

[4] Central tendency is a central or typical value and variability means dispersion around the central value.

Figure 2 RMSE for the NRS Benmore



1.26 Figure 2 shows RMSE for the NRS Benmore. It is clear that RMSE is almost uniformly increasing with time lead. This makes intuitive sense since forecasts should be more accurate closer to the relevant trading period.

Testing pairs of forecasts indicates that the PRS outperforms the NRS

1.27 The next step is to use statistical procedures to conduct pairwise comparisons of forecasting ability. We test whether the PRS has a better forecast accuracy than the NRS and vice versa. The tests point to the overall superiority of the PRS over the NRS. We conclude either that the PRS and the NRS have the same forecasting performance or that the PRS outperforms the NRS. The PRS does better than the NRS in 2-hour forecasts. There are two curious exceptions: the NRS forecasts are better than the PRS forecasts for the first two of 30-minute schedules in Benmore and Haywards.

Performance comparison with ARIMA forecasting: Pre-dispatch schedules consistently perform better

1.28 We tested whether simple forecasting of 5-minute real-time prices gives more accurate information than the DSBF pre-dispatch schedules. The reason for doing this is to compare a simple alternative—an easy-to-estimate time series model of 5-minute prices— to the DSBF schedules to determine whether the DSBF schedules add value to the other existing price forecasts. We investigate this by looking at a year's data starting on the first trading period of 2014. We look at 5-minute real-time prices and take 30-minute averages. We then estimate the most suitable ARIMA model and use it to produce forecasts. The performance of these ARIMA forecasts is compared with those of the 2-hour-ahead schedules of both the NRS and the PRS.

1.29 All the forecasts are show systematic over or under forecasting, particularly in the ARIMA forecasts. All the forecasts under-forecast for prices lower than $100/MWh–$200/MWh and over-forecast for higher prices. There is evidence for non-optimality in all the forecasts. This means that we could make better use of information when producing the forecasts. The ARIMA forecasts show less correlation with observed prices than the NRS and PRS.

1.30 We conclude from the pairwise comparison of the forecasts that both the NRS and the PRS perform better than the ARIMA forecasts. The difference in forecasting performance is statistically significant.

1.31 Model confidence set is a statistical methodology that selects a set of models with superior forecasting ability. The ARIMA forecasts are ranked the least accurate followed by the NRS and the PRS at Otahuhu and Haywards. The NRS is the best performer at Benmore but the ARIMA forecasts are consistently ranked as the worst model. However, none of the models are found to be significantly inferior to the other at any of the nodes. This result is somewhat reflected in the outcome of the measures of forecast accuracy. Both RMSE and the mean absolute errors (MAE) suggest that the NRS and PRS schedules are better than the ARIMA forecasts.

The NRS performs worse than SDPQ; although increased wind generation may explain this result

1.32 We directly compare the forecasting performance of the pre-dispatch schedules before DSBF (the schedule of dispatch prices and quantities SDPQ) with the NRS. This involves statistically testing for the sort of systematic under or over forecasting. There is evidence that all the schedules pre- and post-DSBF of over and under forecasting which gets worse after DSBF. Overall we conclude that the SDPQ is better than the NRS. Note that the NRS includes dispatchable demand.

1.33 Evaluating measures of forecast accuracy and Monte Carlo simulation reveals that precision and uncertainty of the forecasts worsened after DSBF was introduced. This reduction in accuracy could be due to an increase in the amount of installed wind generation over the NRS study period, or it could simply be that load and generation are responding to the price signals that the PRS provides, therefore making the NRS less accurate. In the latter case, DSBF could make the NRS perform worse.

1.34 Another reason for deterioration in accuracy from SDPQ to NRS maybe the change in the timing of load forecast updates and schedule runs. Prior to DSBF, a schedule was run 5 minutes after the load forecasting. Now with NRS, a schedule is run 2 minutes prior to load forecast update and this meant that NRS schedule is 30 minutes behind the SDPQ in incorporating load forecast information. This may make a significant difference at certain times when load is close to the point in the offer stack where price increase rapidly, and the price is very sensitive to the load forecast.

Looking into the behaviour of extreme values

1.35 Since we excluded extreme high prices (prices higher than $2000 / MWh) from our analysis, we isolate them and analyse them separately. The analysis characterises the probability and magnitude of extreme events in the future. We are making inferences about the likely recurrence of these events. A version of a distribution is fitted to the extreme values and the results analysed.

1.36 Extreme values are sometimes more likely to occur in the forecasts than in final prices: more likely for 2-hour forecasts than final prices but less likely for 30-minute forecasts. Extreme values are also more likely to occur in the NRS than in the PRS. The reduced volatility of prices compared to forecasts suggests that participants respond to the pre-dispatch information.

1.37 Extreme values being more likely in the NRS may be attributable to infeasible solutions. This has to do with NRS trying to meet the price inelastic forecast load which may solve with infeasibilities (constraint violation penalty values). The PRS, on the other hand, uses price-dependent bids so can solve at a lower price-quantity solution. Some loads are unmet which reduces total generation back to a feasible level.

Conclusion

1.38 Statistical analysis reveals that DSBF improved the schedule price accuracy of the pre-dispatch process. Including demand response information into the price forecasting process increases the forecast accuracy; the PRS overall performed better than the NRS. As the NRS is effectively the same as the SDPQ we conclude that DSBF is an improvement on the previous regime.

1.39 The forecasts are non-optimal when assessed individually and using multiple horizons. This means that forecasts could be made better by using information in a more efficient way. Price forecast accuracy improves closer to the real time.

1.40 We also find that the SDPQ performed better than the NRS. This could be the result of increased penetration of wind generation; which is intermittent and hence makes price forecasting difficult, or it could be due to generators and load responding to the PRS price signals, therefore making the NRS less accurate. Another reason could be the change in timing of the schedule runs relative to the load forecast—this would affect price forecast accuracy at times when the price was sensitive to the demand forecast. This does not detract from other overall conclusion that the DSBF is an improvement on the previous regime.

Table 1: Summary of changes to price forecasting

<table>
<tr><th colspan="3">Schedules</th><th>Schedule timing</th><th>Non-conforming Load[5]</th><th>Conforming GXPs</th><th>Dispatchable Demand (2014)</th><th>Load forecast</th></tr>
<tr><td rowspan="2">Pre DSBF</td><td>PDS</td><td>Pre-dispatch schedule</td><td>Every 2hours; from 1pm includes the next day too.</td><td>Bids ($ and Q)</td><td>Bids ($ and Q)</td><td>No pre-DD</td><td>N/A</td></tr>
<tr><td>SDPQ</td><td>Schedule of Dispatch Price and Quantities</td><td>every 30 min for current and next 7 trading periods (short)</td><td>Sum of bids (∑ of Q)</td><td>Load forecast</td><td>No pre-DD</td><td>Latest load forecast run 5 min prior to SDPQ</td></tr>
<tr><td rowspan="2">Post DSBF 28-Jun-12</td><td>NRS</td><td>Non-response schedule (Short S and long L versions)</td><td rowspan="2">Every 2 hours for next 36 hours (Long), and every 30 min for current and next 7 trading periods (Short)</td><td>Sum of nominated bids (∑ of Q)</td><td>Load forecast</td><td>Yes from May 2014</td><td rowspan="2">Latest load forecast run 28 min prior to NRS and PRS (Short)[6]</td></tr>
<tr><td>PRS</td><td>Price-response schedule (Short S and long L versions)</td><td>Nominated bids ($ and Q)</td><td>Load forecast +/- any cleared difference bids[7]</td><td>Yes from May 2014</td></tr>
</table>

[5] Prior to DSBF there were simply 'bids', comprising price and quantity pairs indicating the purchasers willingness to buy electricity. Nominated bids have the same structure (price and quantity pairs) pre- and post- DSBF. In the SDPQ and the NRS schedules the price is ignored and only the total quantity is included in the schedule.

[6] The long schedules commence after the latest LF update has taken place. As the long schedules take longer than 1 trading period to complete the results for the first 8 trading periods are not published; if they were they would overwrite more recent short schedule information

[7] Difference bids have been very infrequently submitted. They indicate a deviation from 'normal' in response to the prices signalled in the difference bid. They can signal an intended increment or a decrement.