DEMAND FORECASTING AND PROPOSED INCENTIVES

Authors:

Dr Jakob Blaavand

Dr Rachael Warrington

Dr Robert Leese

Version 2.1

26 January 2017

Smith Institute
Surrey Technology Centre
Surrey Research Park
Guildford GU2 7YG

CONTENTS

1. INTRODUCTION

1.1 This report is the outcome of the contract between the Smith Institute and National Grid to provide support for National Grid's response to Ofgem's consultation "Initial proposals for electricity SO incentives from April 2017".

1.2 The report has two main parts. The first is a commentary of the mathematical underpinnings of Section 4 of Ofgem's proposal, which is Section 2 of this report. The other main part is an investigation of the impact of errors in forecasting solar PV and wind generation on the errors of demand forecasting to give a lower bound for the expected error of the current forecasting models. This is Sections 3 – 7 of this report.

2. COMMENTARY ON PROPOSED OFGEM INCENTIVES

2.1 Ofgem's incentive proposals are laid out in its consultation document and are intended to reduce what it refers to as 'model errors' or 'forecast errors' in the forecasting of wind generation on the timescales of day-ahead, two-days-ahead and week-ahead. Both 'model error' and 'forecast error' are used to refer to the difference between the forecasted value for demand and the actual value that is subsequently recorded. In what follows, we shall use the term 'forecast error'.

2.2 National Grid creates forecasts through a suite of linear regressions. These linear models are constructed using historical actual values for demand and various explanatory factors, most notably various weather variables. They are described in National Grid's Energy Forecasting Training Manual. These models are used to create demand forecasts by taking as inputs forecasts for the various explanatory variables, in particular weather forecasts supplied by the Met Office. The forecast error is determined jointly by any shortcomings in the model itself and by any errors in the input to the model. So, forecast error is some combined effect of model error and input error. Note the different use of 'model error' from the Ofgem proposals, where it is used as a synonym for 'forecast error'.

2.3 The reason for distinguishing between model error and input error is that National Grid is in a much better position to control model error than it is to control input error. It would therefore seem reasonable for National Grid's incentives to be targeted at reducing model error. Model error refers to the errors that result even when the model is supplied with perfectly accurate input data.

2.4 One component in the model error is reflected in the error term of the linear regression, which captures the amount of variation in the historical data that is not explained by the model. Assuming that the historical data is accurate, then the linear regression error term is a model error. Its magnitude is related to the value of R^2 in the regression analysis. As indicated in the Energy Forecasting Training Manual, the values of R^2 in National Grid's models are consistently in excess of 0.95. For each model run, the value of R^2 can be converted to a corresponding contribution to forecast error. The R linear regression package makes the necessary information available. We suggest that it would be useful to look at the regression reports to assess this contribution to the forecast error, which we expect will be rather low.

2.5 Another source of model error arises because of changes in the general energy landscape, which might mean that the historical data on which the models are based, even though accurate, is no longer fully representative of the current electricity system. In other words, there is a contribution to model error because the model is 'past its use-by date'. The recent increase in wind and PV generation over recent months is likely to be a significant contributor to this type of model error. The appropriate action here is to update models more frequently, in line with judgments of when the generation landscape has changed.

2.6 Turning now to input errors, these arise because of estimates (i.e. forecasts) of explanatory variables are used when creating demand forecasts from the models. In general these input values will have errors, which propagate through to errors in the demand forecasts themselves. Even if there are no model errors of the types described above, there will still be input errors. One way of assessing the magnitude of these input errors would be to run the same models again with the benefit of hindsight, i.e. with actual weather data etc., taken on the date for which the forecast was made, in order to see what demand forecast the model would have produced had been no input errors. The comparison of this value with the demand forecast made before the event will show the errors introduced by errors in the input.

2.7 Ofgem proposes four separate incentives: three incentives on forecast accuracy and one incentive for forecast symmetry, i.e. an incentive to produce equal numbers of overforecasts and underforecasts of demand. The errors in linear regression models are naturally symmetric in cases like National Grid’s models, where they do not involve any mathematical transformation of the quantities being forecast.

2.8 The proposed symmetry incentive can easily penalise a perfectly symmetric model, especially through the focus on individual cardinal points. For example, if 30 forecasts are made in a month for particular cardinal point then with probability in excess of 4% there will be either at least 70% overforecasts or at least 70% underforecasts, even if the error in each individual forecast has probability exactly one half of being positive and one half of being negative. Any such event means the maximum penalty is incurred in that month. If there are 10 cardinal points being forecast each day, as in the winter, then the probability of some cardinal point triggering the maximum penalty is more than 35%. So in 35% of winter months the maximum penalty will be applied, even if the model has no error asymmetry at all. In summer months, when there are 12 cardinal points, this rises to more than 40%. In other word, if all forecasts models have perfectly symmetric error terms (which is what the incentive is trying to achieve), then National Grid will incur the full penalty in 4 months every year, on average.[1] In practice, this is likely to happen more often due to input errors. Generally, a small sample of forecasts is more likely to display extreme behaviour than a large sample size.

2.9 The payoff from the proposed symmetry incentive is shown in Figure 1. The maximum reward is only unlocked if exactly half the forecasts are overforecasts and the other half are underforecasts. This makes it very difficult to achieve the maximum reward even if the forecast has a perfectly symmetric error term. In particular, this payoff function will require National Grid to make an even number of forecasts every month to have any chance of the maximum payoff, and not the number of forecasts they find are needed. There are also high

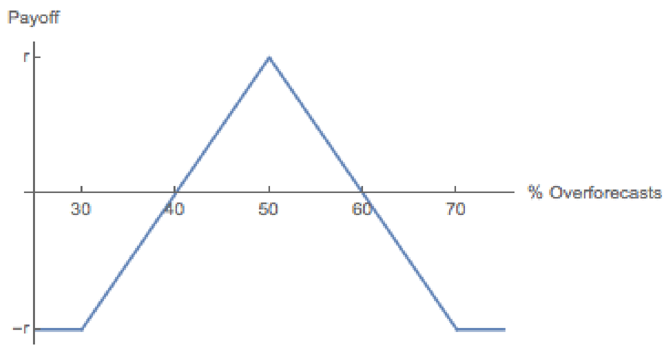

Figure 1 The payoff function suggested in paragraph 4.36 of the Ofgem proposal. The x-axis is the percentage of the forecasts that are overforecasts and the y-axis is the payoff. The maximum reward is denoted r,

[1] These calculations assur … t.

chances of incurring the maximum penalty, especially with small numbers of forecasts and again even if the forecast error is symmetric.

2.10 An alternative payoff function is shown in Figure 2. Here the maximum reward is unlocked if the percentage of overforecasts is between 48% and 52%. With a sample of 300 forecasts made from a model with a perfectly symmetric error term, the maximum reward is achieved about 54% of the times compared to only about 4.5% of the times with the proposed payoff function as shown in Figure 1. The exact numbers and shape of the payoff function are just for illustrative purposes of having an interval that unlocks the maximum reward rather than just a point. In particular, this removes the requirement of having an even number of forecasts every month.

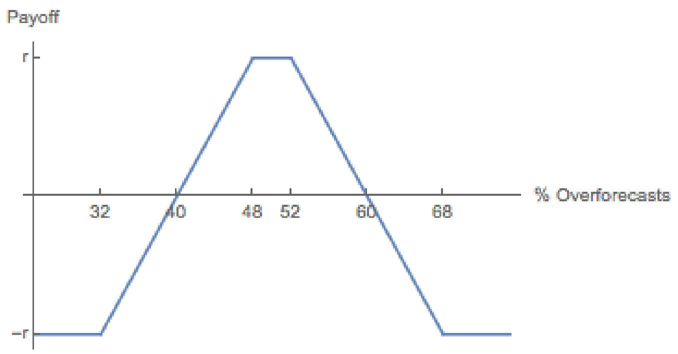

Figure 2 An alternative payoff function for the symmetry incentive. The x-axis is the percentage of the forecasts that are overforecast, and the y-axis is the payoff. The maximum payoff is denoted r. This example retains 40% and 60% with zero payoff, but has changed the points at which the maximum amount is paid from 30% to 32% and 70% to 68% to keep a linear relationship with zero payoff at 40% and 60%.

2.11 The next section looks more closely at the proposals for accuracy incentives. We note that Ofgem does not make clear exactly how performance against the percentage targets for accuracy should be measured, particularly in how the errors over a given month should be averaged. There is also no information of any calculation to support the choice of targets. There seems no justification in particular for why the percentage targets for week-ahead forecasts in Table 4 of the consultation document should be so much lower than for one-day or two-day forecasts. For week-ahead forecasts, one would expect high input errors (it is generally more difficult to forecast weather on longer timescales and during the winter), and therefore less stringent targets. This point is reflected in our analysis in paragraph 6.4 below.

IMPACT OF ERRORS IN FORECASTING PV AND WIND

3. DATA

3.1 The basic dataset for all our analyses is a dataset of demand forecasts from January 2012 to November 2016. The dataset contains forecasts made from a large set of different forecast models and for different forecast horizons. The Ofgem proposal only concerns forecasts made for 1, 2 and 7 days into the future. We calculate the forecast date horizon (in integer days) by taking the difference between the date of the forecasted time and the date the forecast is made. Any negative date horizons, where the forecast was made after the time that it was forecasting for, are neglected. For the remaining forecasts, we calculate the absolute error by taking the absolute value of the difference between the actual demand (settlement demand) and the forecast demand, producing an error distribution for each model in the dataset.

3.2 The data analysed to investigate effects of errors in forecast wind and PV generation are based on a dataset of metered data from all PV installations, on-shore wind installations, and off-shore wind installations that have a capacity of at least 30kW. This dataset contains half hourly meter readings at each site from 31 October 2012 to 31 October 2016. In the analyses these numbers are aggregated across sites to give a national total. The meter data for wind is specified in kW, whereas the meter data for PV is specified in kWh/0.5h. The dataset was shared with us by National Grid.

3.3 Half hourly forecasts for embedded wind and PV generation (in MW) are available from 2012 to 2016 through the National Grid website (http://www2.nationalgrid.com/UK/Industry-information/Electricity-transmission-operational-data/Data-Explorer/). The forecast for embedded generation is the national total including generation from installations of all capacities. We assume that there are very few wind installations with a capacity less than 30kW and that the dataset of metered wind generation is actually the national total wind generated.

3.4 The forecast for embedded PV generation is likewise the national total. Contrary to wind, there are many PV installations with a capacity of less than 30kW, e.g. solar panels on the roofs of individual houses. Indeed, more than 90% of the total national PV capacity up to June 2011 was made up of installations with a capacity of less than 50kW. This breakdown is from Department of Business, Energy & Industrial Strategy[2] and shows the PV deployment in the UK from January 2010 to November 2016. The deployment is broken down into bands of capacity. Unfortunately, there is no break point for 30kW installations, but there are figures for installations with a capacity of at least 50kW. We have used the fraction of 50kW or more capacity to the total PV capacity (as a function of time) to reduce the forecast PV generation to compare it to the values for actual generation in the dataset of metered generation. We will therefore assume a lower forecast figure than what the dataset should show as we are

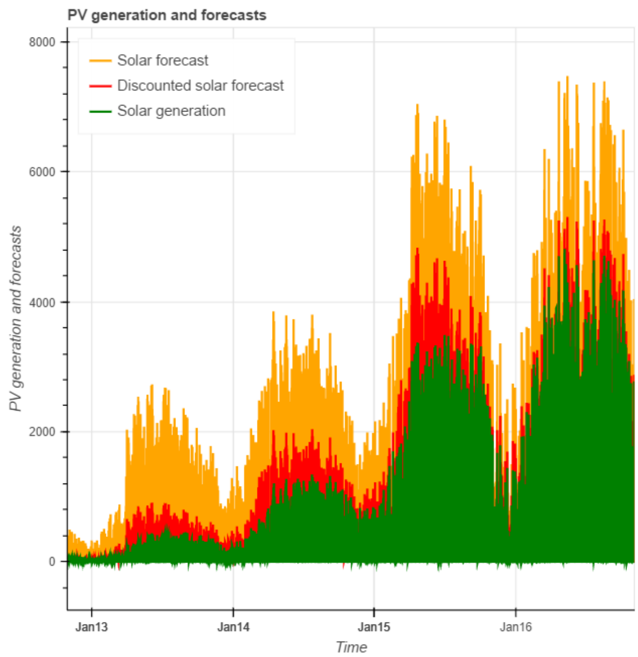

Figure 3 Solar PV forecast generation along with the reduced forecast, where the PV forecast is reduced according to the fraction of capacity of 50kW and above PV installations to the total PV capacity. The green is the total output across all PV sites in the dataset. The x-axis is time, and the y-axis is in units of MW.

[2] https://www.gov.uk/government/statistics/solar-photovoltaics-deployment

excluding sites with a capacity of 30kW – 50kW.

3.5 **Wind and PV error distributions.** The measured wind and PV generation data, together with the forecast wind and PV generation data are used to get error distributions. We calculate, for both wind and solar, the absolute value of the difference between the actual generation and the forecast generation, for each date and time for which both the actual and forecast generation are available. The calculation for wind uses, for each date and time under consideration, the sum of the actual generation for on-shore wind and the actual generation for off-shore wind.

3.6 In Figure 4 the forecast wind and actual generation are plotted along with the signed error of the forecast.

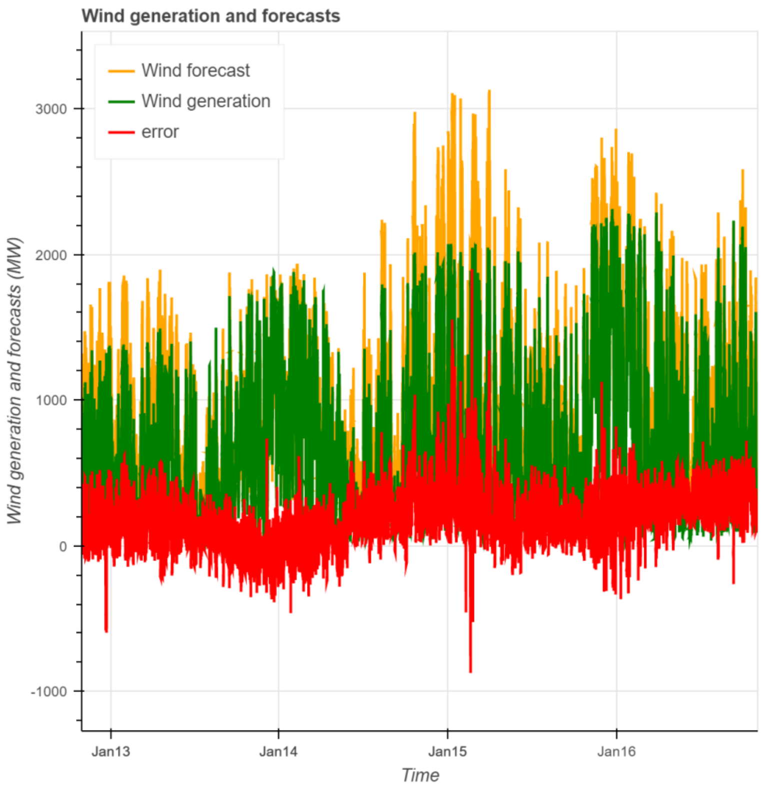

Figure 4 Embedded wind generation forecast and actual wind generated by off-shore and on-shore wind turbines with their signed difference.

4. ASSUMPTIONS

4.1 We assume that the actual generation for on-shore wind and off-shore wind are given in kW, while PV is given in kWh/0.5h, while the forecasts are given in MW. The actual measurements are converted to MW for appropriate calculation of errors. We have not used all the models available in the dataset, but rather have chosen to focus on the cardinal points 1B (morning low), 3B (afternoon low), and 4B / DP (evening high), and on six models that were frequently used to forecast these cardinal points in the years 2012 – 2016. The three specific summer models considered are WKDY1B, WKDY3B, and WKDY4B, and the three winter models considered are GWKDY1B, GWKDY3B, and GWKDYDP. Note that in the proposal we mentioned the winter models GMTDP7 and GMT3B2, but we found that GWKDYDP and GWKDY3B were much more frequently used to predict cardinal points DP and winter 3B, respectively, at the 1, 2, and 7 days forecast horizon.

5. METHODOLOGY

5.1 We consider the error in summer and in winter separately, and for each of three date horizons: 1-day-ahead, 2-days-ahead, and week-ahead. These are the six combinations that Ofgem have proposed targets for, in Table 4 of their consultation "Initial proposals for electricity SO incentives from April 2017". We consider three summer models: WKDY4B7, WKDY1B, WKDY3B7, and three winter models: GWKDYDP, GWKDY1B, GWKDY3B.

5.2 The calculated errors for wind, solar, and national demand are combined and only dates and times for which absolute error values have been calculated for all of wind, solar, and national demand are used. From the dataset of errors, rows corresponding to the model and date horizon of interest are selected, giving a subset of the combined dataset, with each of the six combinations of season and date horizon considered in turn. With each subset of the data, we calculate the mean national demand, together with the mean absolute error and standard deviation of the absolute error for the national demand forecast, forecast wind generation, and forecast solar generation.

5.3 With each of the six subsets of the data, we perform three linear regressions using the statsmodel python package. Each linear regression takes the error in national demand forecasts to be the dependant variable. The first linear regression takes the error for forecast wind generation to be the explanatory variable. The second linear regression takes the error for forecast solar generation to be the explanatory variable. The third linear regression takes both the error for forecast wind generation and the error for forecast solar generation to be the explanatory variables. These linear regressions give R^2-values and p-values. R^2-values tell us how much of the variation in the dependant variable (the absolute error in national demand forecasts) can be explained by a model that is linear in the explanatory variables (the absolute error in the forecast wind and solar generation). The p-values tell us how likely it is that we would get the observed variation in national demand forecast error if the explanatory variable(s) (the absolute error in the forecast wind and/or solar generation) had no effect.

5.4 Another approach would be to consider the relationship between PV, wind, and demand forecasts for each model individually and take a weighted average to get results for the seasonal models. We chose the approach described above as for the week ahead forecasts some of the individual models had too few observations.

6. RESULTS

6.1 In this section we present the results from the regressions described in Section 5.

6.2 Ofgem propose percentage targets for the absolute error in the demand forecast for the three forecast horizons for each summer and winter. The targets are specified in Table 4 of Ofgem's proposal. It is unclear from the proposal how these percentage targets are calculated. Furthermore, when dealing with uncertain quantities, we see it as being inappropriate to give just one percentage target, but rather a confidence interval should be specified for each target.

6.3 Overall, our analyses show that errors in wind and PV forecasts do play a statistically significant role for the errors in forecasting national demand. For each forecast horizon and season the analyses provide an expected MW model error from the current models (along with 95% confidence interval) if the error in PV and wind generation forecasts are zero. Obtaining this level of accuracy for the PV and wind forecasts is, however, highly unlikely, but shows what the inherent forecast errors are when these sources are taken out.

Model	**Horizon**	**PV**	**Wind**	**Multiple**	**PV R^2**	**W R^2**	**Multiple R^2**
Summer	1	TRUE	TRUE	TRUE	0.105	0.051	0.126
Summer	2	TRUE	TRUE	FALSE	0.031	0.002	0.031
Summer	7	TRUE	FALSE	FALSE	0.023	0.003	0.023
Winter	1	TRUE	TRUE	FALSE	0.116	0.002	0.116
Winter	2	TRUE	FALSE	FALSE	0.090	0.001	0.090
Winter	7	TRUE	FALSE	FALSE	0.024	0.002	0.028

6.4 In the table above, the column labelled *PV* indicates whether an error in PV has a statistically significant impact on the absolute error in demand forecasting. Likewise, for the column labelled *Wind*, and the column labelled *Multiple* indicates whether both PV and wind in a multiple regression have a statistically significant impact on the absolute error in demand forecasting.[3] As is clear from the table, most of the explanatory power of the multiple regression comes from the absolute error in PV forecasts.

Model	**Horizon**	**$Mean_{ND}$**	**MAE_{ND}**	**MAE_{ND} (%)**	**MAE_{adj}**	**95% CI**	**MAE_{adj} (%)**	**95% CI**	**# obs**
Summer	1	25,558	466	1.82	312	[288, 336]	1.22	[1.13, 1.31]	3,992
Summer	2	23,995	560	2.33	444	[379, 509]	1.85	[1.58, 2.12]	3,404
Summer	7	26,856	1,249	4.65	921	[537, 1,306]	3.43	[2.00, 4.86]	704
Winter	1	30,676	531	1.73	485	[465, 505]	1.58	[1.52, 1.65]	2,335
Winter	2	29,744	613	2.06	570	[545, 595]	1.92	[1.83, 2.00]	1,872
Winter	7	34,488	1,555	4.51	1,474	[1,341, 1,607]	4.27	[3.89, 4.66]	401

6.5 In the table above, the column labelled $Mean_{ND}$ is the mean of the national demand (in MW) for the observations in the model specified by the first two columns. The column labelled MAE_{ND} is the mean absolute error (in MW) of the forecast demand for all the observations in the model and MAE_{ND} *(%)* is the mean absolute error as a percentage of the mean national demand.

[3] More explicitly, the value is false if one of the variables does not pass the Likelihood Ratio test for reducing the model from two parameters to one. In two cases Wind is statistically significant but taken together with PV it is not. In other words, there seems to be interaction between the two variables.

6.6 The column MAE_{adj} is the adjusted mean absolute error of the demand forecast with respect to the variables that significantly have an influence on the absolute demand forecast error. Only the summer model with forecast horizon of 1 day is adjusted for both PV and wind, all other models are only adjusted for PV. In other words, this column is what the mean absolute error would be if the variable by which it is adjusted for were zero, i.e. if the forecasts of PV and/or wind were perfect. That is, it is the inherent forecasting error in the models, when PV and wind are taken out. The 95% confidence interval (CI) for the adjusted mean absolute error is also given, as well as their conversions into percentages of the national demand.

6.7 The last column is the number of observations in each model.

6.8 The mean absolute errors of national demand given in the table confirm the natural assumption that it is easier to predict the demand tomorrow than two days from now or even a week. This effect should be amplified in the winter as there are larger temperature variations. This is contrary to the targets set in Table 4 of the Ofgem proposal, where the week ahead forecast is proposed to have lower target error than forecasts for the next day.

6.9 It is not clear what a realistically achievable minimum demand forecast error is for each of the six models above. We do not suggest that the adjusted means are realistic targets, but merely that they provide a lower bound for what kind of error could be expected of the models constituting the summer and winter forecasts.

7. LIMITATIONS

7.1 The dataset containing forecasts for wind and PV does not specify when these forecasts are made. Comparing the error distributions between PV, wind, and demand forecast will therefore include errors.

7.2 Despite these limitations we do see a significant effect of errors in PV generation on errors in the demand forecast.

8. CONCLUSIONS

8.1 Regarding the Ofgem incentive proposals, our conclusions are that:

- The averaging involved in performance assessment should be made explicit.
- The symmetry incentive does not seem well designed, and will often impose high penalties on perfectly symmetric models.
- There is no distinction between input error and model error, even though National Grid is in a much better position to influence the latter than the former.
- The proposed incentive targets for week-ahead forecasts are not in line with those on shorter timescales.

8.2 Our investigations show that the absolute error in PV forecasts have a statistically significant impact on the absolute error in demand forecasts for both summer and winter and all time horizons.

8.3 Adjusted means of absolute errors in demand forecasting for each model are given in the table in paragraph 6.4 and provide a lower bound for the error that can be expected of the current models constituting the summer and winter forecasts.