James W. Taylor **29 April 2023**

External Examiner's Report on the Doctoral Thesis of Bartosz Uniejewski of Wroclaw University

Thesis title: Forecasting wholesale electricity prices to support decision-making in power companies: Use of regularization and forecast combinations

This thesis is concerned with short-term forecasting of electricity prices. The first part of the thesis provides background to the area and an overview of the five papers that constitute the substantive part of the thesis. The remainder of the thesis consists of the five papers. In the rest of this report, I first discuss each of the five papers. For each, I give a description of its contents and my view of the paper. At the end of this report, I summarise my overall views on the thesis, and provide appendices with thoughts on potential future work and minor comments.

Paper 1

Paper 1 is co-authored with Grzegorz Marcjasz and Rafał Weron, and was published in the *International Journal of Forecasting* in 2019.

The paper concerns point forecasting of electricity prices in an intraday market. The motivation provided for this is that intraday markets are becoming increasingly important due to the expansion of renewable generation, modernization of power grids, and increased demand-side management.

The paper was the first to study the forecasting of prices for the German EPEX intraday market. Only two previous studies had been written on forecasting prices in European intraday markets, and in both, variable selection was important due to there being a large number of potential regressors. In Paper 1, this situation is recreated with regression models that have, as regressors, many different lags of the intraday price, many different past, current and future forward-looking day-ahead prices, and weekday dummy variables.

To address the issue of variable selection, Paper 1 evaluates the usefulness of the least absolute shrinkage and selection operator (LASSO). LASSO involves the estimation of a regression model with the inclusion of a penalty term in the parameter optimisation, which has the effect of limiting the number of none-zero coefficients in the regression model.

In addition to bringing LASSO to the intraday price forecasting context, a strength of the paper is the thorough evaluation of the point forecasts. Alongside the common statistical error measures, Diebold-Mariano tests are used to compare the statistical error measures, as well as a calculation of trading profit. Also, the models are compared for the out-of-sample period divided up into periods with positive spikes, negative spikes, and no spikes, and also divided up according to the day of the week.

The paper provides a very good start to the set of five papers. Its quality is evident from it winning the award for best energy forecasting paper published in the *International Journal of Forecasting* in the years 2019- 2020. This is an impressive achievement, as this is the leading journal in forecasting.

Paper 2

Paper 2 is co-authored with Katarzyna Maciejowska, and has been accepted for publication in the *International Journal of Forecasting*.

The paper concerns day-ahead point forecasting of electricity prices for each hour of the next day. Data from four different day-ahead markets are considered. The regressors included in the regression models differ a little across the four markets, and include lags of price, load forecasts, wind and solar generation forecasts. A particular transformation is applied that converts the unconditional distribution of the prices into a Gaussian distribution.

The fundamental issue that this paper addresses is the choice of the length of the estimation window to use when estimating the parameters for a regression model. A shorter window has the appeal of focusing on recent data, which is useful for rapidly evolving markets, but it has the downside of increased sampling error and, as the authors point out, limits the potential for considering a complex model. This paper follows the suggestion of previous studies that it can be useful to estimate models using multiple different estimation sample lengths and then combine the forecasts produced by the various models. The authors consider many different estimation sample lengths, ranging from 56 to 728 (i.e., 673 models), meaning that they need to combine a panel of many forecasts (i.e., 673 forecasts).

A natural way to combine the forecasts is to use them as regressors in a regression model. The authors highlight two approaches proposed in the literature to address the large number of regressors. The first is the use of LASSO. The second is termed principal component averaging, and applies principal component analysis (PCA) to the panel of forecasts to reduce the number of regressors, and then uses the principal components (PCs) as regressors.

The new approach proposed in Paper 2 is a synthesis of these two approaches. PCA is applied to the many forecasts, and then the principal components are used as regressors in LASSO with the tuning parameter chosen using an information criteria. The authors term this *LASSO principal component averaging (LPCA)*. The advantage of LPCA over the simpler use of LASSO is that the PCA helps synthesise the information available in the many forecasts, which simplifies the task for the LASSO. The authors also highlight computational advantages of LPCA over LASSO. The advantage of LPCA over the simpler PCA approach is that, when a number k of PCs are chosen for the PCA approach, all PCs from 1 to k are used, while the LPCA approach enables any set of PCs to be selected.

An empirical study compares the new approach to relatively simplistic combinations, as well as the combination approaches suggested in the literature based on LASSO or PCA. Overall, the results are the best for the new LPCA approach.

In my view, the proposed new approach is a very nice amalgamation of the simpler LASSO and PCA approaches to combining a large number of point forecasts. I find the motivation for the method persuasive and the results encouraging.

Paper 3

Paper 3 was co-authored with Grzegorz Marcjasz and Rafał Weron, and was published in *Energy Economics* in 2019.

The paper concerns day-ahead probabilistic price forecasting. Unlike Papers 1, 2 and 4, Paper 3 involves no regularization methods.

This paper builds on the work of Nowotarski and Weron (2016), which showed that price point forecast accuracy benefits when the price series is decomposed into a long-term seasonal component (LTSC) and a stochastic residual component, an ARX model is fitted to each component to produce a forecast for each, which are then summed to give a forecast for price. While previous work considered only point forecasting, the current paper evaluates the use of the approach as the basis for probabilistic forecasting. A further aim of Paper 3 is to evaluate whether the decomposed components are each best predicted by combining the component predictions from different models.

Using data from two markets, the empirical study involved the following: (i) To decompose the (log) price series, two approaches are used: Hodrick-Prescott filter and wavelet smoothing. Using different filtering and smoothing parameters, led to 54 different decompositions. (ii) To model the LTSC and residual component, ARX models are used. The authors find benefit in applying the decomposition approaches to regressors, such as electricity load, and then using the resulting LTSC as regressor when modelling the LTSC for price. (iii) To produce distributional forecasts, three methods are used: historical simulation; bootstrapping, which accounts for parameter uncertainty as well as residual uncertainty; and quantile regression averaging (QRA), which constructs a distribution function from quantile forecasts produced using quantile regression with regressors chosen as a set of point forecasts from multiple models. (iv) To combine distributional forecasts, two methods are used: probability averaging and quantile averaging.

Distributional forecasts are evaluated using the average of the quantile score for the 99 percentiles from 1% to 99%, and using the average of the quantile score for the lowest 10 percentiles and highest 10 percentiles. Significant superiority of one method over another is tested using the Diebold-Mariano test. The results are encouraging for the LTSC approach with ARX modelling of the components with distribution constructed using QRA. In terms of distributional forecast combining, they find that probability averaging is more accurate than quantile averaging. This is of particular interest to me as some of my research is in the area of distributional forecast combining.

Paper 3 makes a useful contribution to the literature in extending the previous work of Nowotarski and Weron (2016), regarding LTSC decomposition to the probabilistic forecasting context.

Paper 4

Paper 4 was co-authored with Rafał Weron, and published in *Energy Economics* in 2021.

This paper concerns day-ahead probabilistic price forecasting. The contribution of the paper is a new version of QAR that involves automated regressor selection via the use of LASSO regularization. They abbreviate the new method as *LQRA*. Motivation for this is the empirical observation that QAR can perform poorly when there is a large number of point forecasts in the combination.

The data used in the empirical analysis is for the Polish and Nordic power markets. A variance stabilizing transformation was initially applied to the price series with the main aim to reduce the impact of price spikes. Standardisation is also applied.

Point forecasts are produced using an ARX model fitted separately for each hour of the day. The regressors include autoregressive terms, transformed day-ahead load forecasts, transformed forecasts of generation, and day-of-the-week dummy variables. Multiple point forecasts are created by estimating the ARX model using estimation windows of length between 28 to 728 days, with increments of 28 days, which leads to 25 different window lengths and hence 25 point forecasts.

The benchmark methods include QRA, the quantile regression machine of Uniejewski et al. (2019), as well as distributional forecast combining via quantile or probability averaging. The methods are compared in terms of their ability to forecast the 50% and 90% intervals. For this, the interval score and interval calibration are used. The test of Giacomini and White (2006) is used to test for significant difference between the scores for pairs of methods, and the Kupiec test is used to evaluate unconditional calibration. Evaluation is also performed in terms of the profit resulting from a trading strategy based on the high and low point of prices on the next day.

The proposed new LQRA method performs well relative to the other methods, with better results achieved with the use of BIC, rather than cross-validation, to optimise the LASSO tuning parameter.

I very much like this paper. Although the QAR approach has performed well in empirical studies, a weakness has been identified, and in Paper 4, the authors address this through the incorporation of LASSO in the approach. This seems to me a sensible innovation, and the results are encouraging.

Paper 5

Paper 5 is co-authored with Katarzyna Maciejowska and Rafał Weron, and has been accepted for publication as a chapter in *The Oxford Research Encyclopedia of Economics and Finance*.

After first describing the structure of day-ahead and intraday markets, this paper focuses on three areas where the authors feel there has been, and are, significant developments. The first is the rapidly increasing interest in probabilistic rather than point forecasts. The rising interest in this reflects a trend seen in the forecasting literature over the past couple of decades. The second area, the authors describe as seeing significant development, is the use of complex statistical learning/machine learning methods in preference to parsimonious statistical methods. This also reflects a trend in the broader forecasting literature. The third area considered in the review is the evaluation of forecast accuracy in terms of financial benefit, to complement more traditional statistical measures of accuracy.

A very nice aspect of this review paper is that, rather than simply describe the many forecasting methods that have been proposed in the literature, the authors provide their own perspective on three quite distinct areas that are receiving considerable attention. It is likely to be a useful resource for newcomers to short-term electricity price forecasting, as well as for researchers wanting an update of recent developments. As with many review articles, it is likely to be well cited, particularly, as noted by the authors, there are very few review papers on this topic.

Summary

In my view, the thesis is very nicely organised and presented, with helpful illustrative figures. The writing is of a very high standard, making the thesis easy to read. Importantly, the clarity of explanations enables any empirical experiments to be reproduced by authors. For example, there are clear explanations for the methods considered, and the sample periods used for estimation, cross-validation and evaluation.

In terms of novel contributions, the thesis contains several interesting new forecasting methods, provides thorough empirical evaluations of a variety of forecasting methods, and presents a very useful review of the research area. All five papers are of high quality, which has been demonstrated by them being published or accepted for publication in respected international journals and, in the case of Paper 5, as a book chapter. I particularly like Papers 2 and 4. Paper 2 merges LASSO and PCA approaches to improve the combining of a large set of point forecasts, while Paper 4 incorporates LASSO within the popular and successful QRA method to improve probabilistic forecast accuracy.

Appendix 1 lists general issues of interest to me, and Appendix 2 provides minor comments. None of the issues that I have raised could be classed as major concerns. I certainly feel the candidate should be awarded the doctoral degree.

Signed:

James W. Taylor

Professor James W. Taylor
Saïd Business School
University of Oxford
Park End Street
Oxford OX1 1HB, UK

Appendix 1 – Comments on Potential Future Work

1. In Papers 1, 2 and 4, I was not clear as to whether standardisation was performed on the regressors prior to the use of LASSO. I would have thought this would be useful.

2. It is interesting that, in the literature on probabilistic price forecasting, distributional forecasting has often been performed by separately modelling the 99 percentiles. With 99 such models for each of the 24 hours of the day, a large number of models are needed. One has to wonder whether a different approach could be used that models different percentiles and hours simultaneously, with the potential aims being to capture the information more efficiently and ensure coherent distributional forecasts (i.e., non-crossing quantiles).

3. The thesis does not mention the use of generalised autoregressive conditional heteroscedasticity (GARCH) models to try to model time-varying volatility in transformed price series. I appreciate that the variance-stabilising transformation will have an impact on the volatility structure, but I still wonder whether there may be remaining volatility dynamics that could be modelled. Another possibility is autoregressive quantile modelling.

4. My understanding is that the literature shows that quantile regression combinations of a set of quantile forecasts from multiple models is outperformed by QRA, which involves quantile regression with regressors chosen as a set of forecasts of the mean produced by multiple models. In a sense, for a distribution with time-varying mean, it is not surprising that having good forecasts of the mean of a distribution is a good basis on which to forecast that distribution. However, if there are dynamics in the distribution's higher moments that are distinct from those governing the mean, only having good forecasts for the mean would seem to be a limited basis for forecasting the higher moments. If the distribution has different dynamics in different parts of the distribution, and one could, to an extent, capture this in the set of multiple forecasting models, then one would assume the quantile forecasts from such models would be preferable to forecasts of the mean when choosing regressors to include in a quantile regression model. However, if the multiple quantile forecasts are based on the assumption that the only dynamics in the conditional distribution are in the mean, then such quantile forecasts would be no better than multiple forecasts of the mean, when choosing regressors to include in a quantile regression model.

5. The introduction to Paper 4 cites value-at-risk (VaR) as a motivation for probabilistic price forecasting. However, the financial risk management literature highlights the inadequacies of VaR as a risk measure, and now recommends expected shortfall (ES). Therefore, I wonder whether ES forecasting should be considered for electricity prices, rather than just VaR. Of course, forecasting the 99 percentiles would not be sufficient to obtain a forecast for the ES.

6. Paper 5 briefly discusses the forecasting of ensembles, which are also known as paths or trajectories. It would be interesting to consider how the methods described in the paper could be adapted to forecast, not just the marginal distribution for each hour of the day, but also the interdependence between the hours.

Appendix 2 – Minor Comments

Paper 1 – On page 1536, it is stated that the idea underlying a variance stabilizing transformation is to reduce the price variation. But this is not quite right. Instead, the aims can be to make the variance not depend, or depend less, on the mean, and/or to make the distribution more symmetric and closer to Gaussian. I note that this point is recognised in the other papers.

Paper 1 – Why is the area hyperbolic sine transformation used?

Paper 2 – Would there be any worth in also using PCA to reduce the problem from forecasting the 24 hours of the day to forecasting a lower number of principal components?

Paper 2 – In the conclusions, it is stated that the method could be extended for probabilistic forecasting. How?

Paper 2 – I wonder whether the positive results for LASSO are sensitive to the choice of the set of models used (which all have the same regressors).

Paper 3 – How is quantile crossing handled in the QRA approach?

Paper 3 – Are the tails of the distribution, and in particular the spikes, sufficiently well captured by the 1% and 99% quantiles? In the QRA and in the score used for evaluation, would there be benefit in using more extreme quantiles? Perhaps it would be beneficial to use more quantiles in the tails and fewer in the centre of the distribution.

Paper 3 – Is it wise to use only the n=1, 2, …, 19 best performing distributional forecasts in the combinations? Perhaps the quantile averaging method would have been more successful if there had been greater variation among the locations of the distributional forecasts.

Paper 3 – In this paper, probability averaging outperforms quantile averaging for distributional forecast combining. In view of this, when considering the construction of an individual distributional forecast, instead of QRA, perhaps probability model averaging could be used, where a separate logistic autoregression is used for each of a set of values for the log price.

Paper 3 – It would also have been interesting to have had calibration results using quantile coverage or PIT histograms.

Paper 4 – With regard to the choice of LASSO tuning parameter, it seems natural for the optimal value to vary with the quantile probability level α, as considered in the paper. But perhaps a single tuning parameter could be used for all values of α, if, in the penalty term, it is multiplied by an approximation of the quantile regression loss function. The approximation could be based on an initial quantile regression performed with no LASSO penalty term.

Paper 4 – It seems a little disappointing that LASSO does not perform well when selecting from the point forecasts produced by models for all 701 window lengths between 28 to 728 days. Using just the 25 selected window lengths seems rather arbitrary.

Paper 4 – Is it reasonable to deal with quantile crossing by sorting the 99 quantile forecasts? Perhaps there is not much quantile crossing.

Paper 4 – Section 4.2.4 is entitled Sharpness, but then discusses the quantile score averaged across all 99 percentiles. But this is a proper score for a distributional forecast and so it is a measure of both calibration and sharpness. I note that this is recognised in the review of Paper 5.

Paper 4 – Conditional calibration could also have been evaluated. I note that it is mentioned in the review of Paper 5.