ARTICLE IN PRESS

International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging

Katarzyna Maciejowska [a,b], Jakub Nowotarski [a], Rafał Weron [a,*]

[a] Department of Operations Research, Wrocław University of Technology, Wrocław, Poland

[b] CERGE-EI, Prague, Czech Republic

ARTICLE INFO

Keywords:
Probabilistic forecasting
Prediction interval
Quantile regression
Factor model
Forecasts combination
Electricity spot price

ABSTRACT

We examine possible accuracy gains from using factor models, quantile regression and forecast averaging to compute interval forecasts of electricity spot prices. We extend the Quantile Regression Averaging (QRA) approach of Nowotarski and Weron (2014a), and use principal component analysis to automate the process of selecting from among a large set of individual forecasting models that are available for averaging. We show that the resulting Factor Quantile Regression Averaging (FQRA) approach performs very well for price (and load) data from the British power market. In terms of unconditional coverage, conditional coverage and the Winkler score, we find the FQRA-implied prediction intervals to be more accurate than those of either the benchmark ARX model or the QRA approach.

1. Introduction

Over the last two decades, electricity spot price forecasting has become the core process of an energy company's planning activities at the operational level. Statistical/econometric approaches (like multiple regressions, AR, ARIMA, AR-GARCH, jump-diffusions, factor models and regime-switching models) and computational intelligence techniques (like neural networks, fuzzy techniques and support vector machines) constitute the two main streams of models, both in the academic literature and in actual business practice (see e.g. Amjady & Hemmati, 2006; Chan et al., 2012; Hong, 2014; Weron, 2006, 2014; Zareipour, 2008).

While there have been a variety of empirical studies on point forecasts (i.e., the 'best guess' or expected value of the spot price), probabilistic – i.e., interval and density – forecasts have not been investigated extensively to date (for a recent comprehensive review, see Weron, 2014). Such is the case in spite of the fact that, from a practical point of view, prediction intervals (PI), and density forecasts even more so, provide additional information on the evolution of future prices. In particular, Chatfield (2000) lists the following reasons for the importance of interval forecasts: (i) the assessment of future uncertainty; (ii) an ability to plan different strategies for the range of possible outcomes indicated by the interval forecast; and (iii) the possibility of more thorough forecast comparisons. Also, electrical engineers are aware that high-quality market clearing price PI would help utilities to submit effective bids with low risks (Amjady & Hemmati, 2006). Nevertheless, the literature on the interval forecasting of electricity prices is very sparse, probably due to the increased complexity of the problem compared to point forecasting. The fact that some authors use the term 'confidence interval' instead of 'prediction interval' (PI) adds confusion. However, in most forecasting applications we are interested in PI associated with a random variable (e.g., the electricity price) that is yet to be observed, i.e., intervals which contain the true values of future observations with a specified

* Corresponding author.
E-mail addresses: katarzyna.maciejowska@pwr.edu.pl (K. Maciejowska), jakub.nowotarski@gmail.com (J. Nowotarski), rafal.weron@pwr.edu.pl (R. Weron).

http://dx.doi.org/10.1016/j.ijforecast.2014.12.004
0169-2070/

probability, rather than in confidence intervals that are associated with a parameter of a model; see Hyndman (2013) for a discussion.

The second issue that bothers electricity price forecasters is the fact that, given the diversity of models, it is essentially impossible to select *ex ante* a single, most reliable one. In this context, combining forecasts has the potential to turn out to be particularly useful. Indeed, as Bordignon, Bunn, Lisi, and Nan (2013), Nowotarski, Raviv, Trück, and Weron (2014) and Raviv, Bouwman, and van Dijk (2013) report, combining does lead to more accurate (on average) and more robust electricity price point forecasts.

What about probabilistic forecasts though? Although the idea of combining interval forecasts in itself is not new (Timmermann, 2006; Wallis, 2005), it has not previously been utilized in the context of electricity spot prices. However, Nowotarski and Weron (2015) recently introduced a novel method of computing prediction intervals (PI) and dubbed it Quantile Regression Averaging (QRA; see Section 4.1 for a brief account). The method involves applying quantile regression to a pool of point forecasts of individual (i.e., not combined) forecasting models. Using PJM market data, Nowotarski and Weron (2015) showed the QRA-implied PI to be more accurate than those obtained using any of the 12 individual time series models considered, in terms of both unconditional and conditional coverages. In a parellel study Nowotarski and Weron (2014) evaluated the QRA method further, and, using Nord Pool day-ahead prices, provided even more convincing evidence in favor of the new approach. Our aim here is to extend their approach and to address more efficiently both points that have been mentioned as issues, namely (i) selecting a set of models that performs well for combining and (ii) constructing accurate prediction intervals. To this end, we use principal components to automate the selection process from among a large set of point forecasts of electricity spot prices; this is in contrast to Nowotarski and Weron (2014, 2015), who select a set of 6 (12) individual models *a priori*. Then, as in these two papers, we apply quantile regression, but to point forecasts of the principal components (i.e., the common factors) rather than to point forecasts of the individual models.

The remainder of the paper is structured as follows. In Section 2, we briefly present the British power market dataset studied here. Section 3 describes the 32 individual models used for obtaining point forecasts of electricity spot prices, after which Section 4 reviews the Quantile Regression Averaging (QRA) method of Nowotarski and Weron (2015) and introduces a novel method for computing prediction intervals, called Factor Quantile Regression Averaging (FQRA). Section 5 evaluates the forecasting performances of the three approaches tested (the benchmark ARX model, QRA and FQRA) in terms of their unconditional coverages, conditional coverages and Winkler scores (also known as the interval score). Finally, Section 6 wraps up the results and concludes.

2. The data

The dataset used in this study comprises British volume-weighted reference prices for each half-hourly load period, together with half-hourly day-ahead load forecasts, both for the period July 1st, 2010–December 31st, 2012; see Fig. 1. The prices were obtained from the APX power exchange (www.apxgroup.com; note that volume-weighting is performed over three types of contracts: half-hourly, two-hour-block and four-hour-block contracts), and the load forecasts – i.e., the National Transmission System Demand Forecasts – from the system operator (www.bmreports.com). Both series have been preprocessed for missing and 'doubled' values (due to changes to/from daylight saving times) in the traditional way, see for example Weron (2006). The logarithms of half-hourly load forecasts are used as the exogenous variable in the time series models for the log-prices. This selection is motivated by a roughly linear dependence between these two variables.

The dataset is split into three subsets. The first date given, July 1st, 2010, marks the start of the 365-day-long calibration period for the 32 individual models (for model definitions, see Section 3). The first day-ahead predictions of these models are obtained for the 48 half-hourly load periods of July 1st, 2011, the second date in Fig. 1. Then, the calibration window is rolled one day forward, the individual models are recalibrated, and spot price predictions are made for July 2nd, 2011, etc. Finally, the third date in Fig. 1, January 1st, 2012, indicates the first day for which interval forecasts were obtained by applying quantile regression either to point forecasts of the individual models (for the QRA approach; see Section 4.1) or to factors calibrated to those forecasts (for the FQRA approach; see Section 4.2). For the latter, a rolling window of a fixed length (184 days or roughly half a year; initially, from July 1st to December 31st, 2011) is also used. The interval forecast validation period lasts until December 31, 2012, and includes 366 days.

3. Individual models

A typical and obvious feature that is shared by all empirical applications that use forecast averaging is that the results depend on the specific choice of individual models. Thus, our selection of models includes a set of carefully picked model classes that have been applied in the energy economics literature: autoregressive models (AR-type), threshold autoregressive models (TAR-type), semiparametric autoregressive models (SNAR-type), non-linear AR (neural network) models (NAR-type), and principal component (or factor) models (PC-type). All of the individual models have an underlying autoregressive structure. Here, two sets of lags are considered: either all lags from 1 to 7 are included (as per Maciejowska & Weron, 2013), or only lags 1, 2 and 7 are used (as per Weron & Misiorek, 2008). For a full list of models, see Table 1. Furthermore, in all models except for PC-type models, three types of deterministic variables are included: a constant, a weekend dummy (taking the value of 1 for Saturday and Sunday and 0 otherwise; a proxy for weekly seasonality), and day length (from a sunrise to a sunset; a proxy for annual seasonality).

3.1. Autoregressive models

The general autoregressive structure underlying all of the individual models in this study has the following