HSC Research Report

HSC/15/04

Short- and mid-term forecasting of baseload electricity prices in the UK: The impact of intra-day price relationships and market fundamentals

Katarzyna Maciejowska[1]
Rafał Weron[1]

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

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

IEEE Transactions on Power Systems,
http://dx.doi.org/10.1109/TPWRS.2015.2416433

Short- and mid-term forecasting of baseload electricity prices in the UK: The impact of intra-day price relationships and market fundamentals

Katarzyna Maciejowska and Rafał Weron

***Abstract*—In this paper we investigate whether considering the fine structure of half-hourly electricity prices, the market closing prices of fundamentals (natural gas, coal and CO_2) and the system-wide demand can lead to significantly more accurate short- and mid-term forecasts of APX UK baseload prices. We evaluate the predictive accuracy of a number of univariate and multivariate time series models over a three-year out-of-sample forecasting period and compare it against that of a benchmark autoregressive model.**

We find that in the short-term, up to a few business days ahead, a disaggregated model which independently predicts the intra-day prices and then takes their average to yield baseload price forecasts is the best performer. However, in the mid-term, factor models which explore the correlation structure of intra-day prices lead to significantly (as measured by the Diebold-Mariano test) better baseload price forecasts. At the same time, we observe that the inclusion of fundamental variables – especially natural gas prices (in the short-term) and coal prices (in the mid-term) – provides significant gains. The CO_2 prices, on the other hand, generally do not improve the price forecasts at all, at least in the time period considered in this study (Apr. 2009 – Dec. 2013).

***Index Terms*—Electricity price, Forecasting, Vector autoregression, Factor model, Principal components**

I. Introduction

OVER the last 15 years electricity price forecasting has become the backbone of an energy company's decision-making process [1]. Short-term (from a few minutes up to a few days ahead) and mid-term (up to a few months ahead) price forecasts have become of particular interest to power portfolio managers. The short-term forecasts of the intra-day prices (typically hourly or half-hourly) are of prime importance in day-to-day market operations, in particular when bidding at a power exchange or for implementing effective demand response [2]–[5]. The mid-term forecasts, on the other hand, are generally used for planning purposes (such as the adjustment of mid-term schedules and allocation of resources), risk management (including balance sheet calculations) and the valuation of exchange traded futures and bilateral contracts [6]–[9]. In most cases, these tasks concern the baseload electricity price, i.e. the average price for the 24 hours of the day, or the peakload price, i.e. the average price for the peak hours. In particular, the most common underlying instrument of exchange traded power derivatives is the baseload price. Hence, like Garcia-Martos et al. [10], we focus in this study on forecasting baseload electricity prices.

The literature on predicting baseload electricity prices has concentrated on models that use the available price data only at the aggregated (i.e. daily) level, see e.g. [10]–[14]. On the other hand, the very rich literature on forecasting intra-day prices has used hourly or half-hourly data, but generally has not explored the complex dependence structure of the multivariate price series. A notable exception is a working paper of Wolak from 1997, published as [15], in which principal component analysis (PCA) is applied to hourly or half-hourly prices from the UK, Scandinavia, Australia and New Zealand, to understand the price formation mechanism and measure the relative predictability of the daily vector of prices in each country.

Only a decade later has the multivariate context of electricity prices been picked up again. Chen et al. [16] use manifold learning (an extension of principal component analysis, PCA) to remove intra-day and intra-week seasonality from hourly electricity prices and predict them using three techniques. Their approach compares favorably to that of ARIMA, ARX and naive methods in one day, one week and one month ahead forecasting of hourly NYISO prices. Härdle and Trück [17] utilize dynamic semiparametric factor models (DSFM) to forecast hourly electricity prices in the German EEX market. They find that a model with three factors is able to explain up to 80% of the variation in hourly prices, however, the explanatory power significantly decreases for periods with a higher number of price spikes. Alonso et al. [18] develop the Seasonal Dynamic Factor Analysis (SeaDFA) to deal with dimensionality reduction in such a way that both common and specific components are extracted. This approach is applied to compute long-term (up to a year) point forecasts and prediction intervals of electricity spot prices. However, as Garcia-Martos et al. [19] argue, SeaDFA is very sensitive to the choice of the Vector ARIMA (i.e. VARIMA) model for the common factors and underperforms if a wrong model is selected. As a robust alternative they propose the dynamic factor model (DFM) framework to extract common factors from hourly prices and use them for one day-ahead forecasting. They also report some preliminary results showing the usefulness of factor models for mid-term predictions. Wu et al. [20] introduce a recursive dynamic factor analysis (RDFA) algorithm and show that it

This work was partially supported by the Ministry of Science and Higher Education (MNiSW, Poland) core funding for statutory R&D activities and by the Croatian Science Foundation under grant no. IP-2013-11-2203.

KM is with the Department of Operations Research, Wrocław University of Technology, Wyb. Wyspianskiego 27, 50-370 Wrocław, Poland and CERGE-EI, Politickych veznu 7, 11121 Prague 1, Czech Republic. RW is with the Department of Operations Research, Wrocław University of Technology, Wyb. Wyspianskiego 27, 50-370 Wrocław, Poland. (e-mails: {katarzyna.maciejowska, rafal.weron}@pwr.edu.pl).

outperforms functional PCA, AR with time varying mean and support vector regression (SVR) in predicting hourly day-ahead prices in the Australian and New England markets.

Finally, in an article that is the most similar to our study, Garcia-Martos et al. [10] build a VARIMA(0,1,1) model for baseload electricity and daily fossil fuel prices (oil, natural gas and coal; plus CO_2 prices). Then they compare its forecasting performance to that of univariate benchmark models (ARIMA) for each daily series. They consider a range of forecasting horizons (from one to 15 business days) but evaluate the predictions only over a relatively short test period (the first 4 months in 2011). Overall they find that for the electricity (and CO_2) prices the univariate approach produces better results, whereas for oil, natural gas and coal prices the multivariate approach is more accurate. Additional inclusion of wind power generation improves the baseload electricity price forecasts, but only for the one day-ahead horizon. In the later part of the article, which goes in a different direction than our study, Garcia-Martos et al. extract the common features in volatility by means of a conditionally heteroskedastic DFM and claim that the obtained common volatility factors are useful for improving the quality of prediction intervals (PI). However, the obtained PI are not tested for coverage, as suggested e.g. in [1], only a visual inspection of the conditional volatility is performed.

In this study, we pursue a similar task and attempt to improve the accuracy of baseload electricity price forecasts in the short- and mid-term horizons. Like Garcia-Martos et al. [10], we work with daily time series and consider fossil fuel prices. However, unlike them, we additionally analyze the information embedded in the intra-day electricity price relationships and use it to provide more accurate predictions of baseload prices. Note also that their notion of a 'multivariate model' is different than ours. They use electricity and fossil fuel prices sampled at daily frequency and model them jointly within vector autoregression type models, while we regard as multivariate only those models which utilize the information contained in the fine structure of half-hourly electricity prices and treat the daily closing prices of fundamentals as exogenous variables.

While new to the electricity price forecasting literature, the idea of using disaggregated data for forecasting of aggregated variables has been exploited in the economic literature, particularly in macroeconometrics – to predict inflation [21], [22], the Gross Domestic Product [23] or the Production Index [24]. See also Lütkepohl [25], who warns that the inclusion of too many disaggregates can result in estimation error and specification error which ultimately leads to an efficiency loss, and Hendry and Hubrich [26], who describe conditions under which using disaggregated data improves forecasting performance.

In the context of electricity markets this concept has emerged only very recently. The very few publications include Liebl [27], who suggests to model and predict baseload prices by first finding the functional relation between electricity prices and demand in terms of daily price-demand functions, then parametrizing the series of daily price-demand functions using a functional factor model. He demonstrates the power of this approach by comparing one to 20 step-ahead baseload price forecasts of the model with those of two simple univariate time series models for baseload prices (AR and MRS) and two alternative functional data models for hourly prices (DSFM and SFPL). In a limited empirical study, Maciejowska and Weron [28] use half-hourly data from the UK power market to forecast the baseload spot prices directly (via AR and vector AR models) and indirectly (via factor models). The results indicate that there are forecast improvements from incorporating the disaggregated (i.e. half-hourly) data, especially, when the forecast horizon exceeds one week. Raviv et al. [29] exploit the information embedded in the cross correlation of Nord Pool hourly price series to yield more accurate day-ahead baseload price forecasts.

In this paper, we extend the above mentioned studies in several directions. Firstly, we analyze a wider range of forecasting horizons – from one to 45 business days (or working days, i.e. excluding weekends and holidays; the far end corresponds to a little over two calendar months). Secondly, we consider more diverse model structures and different aggregation levels. Initially we have also used models where the fundamental variables were predicted jointly with the prices (or factors), within vector autoregressive (VAR) structures. However, such models – similar to the VARIMA(0,1,1) models considered in [10] – have turned out to yield inferior forecasts compared to models where the fundamental variables are treated as exogenous variables and predicted independently. Consequently, we have decided to focus in this article only on the latter. Thirdly, the influence of including various fundamental variables on the predictive performance of the models is studied. So far, there have been no publications, which discuss both the optimal choice of the variables and the level of aggregation. Since mid-term forecasts play a crucial role in planning activities (such as the adjustment of mid-term schedules and allocation of resources), risk management and the valuation of exchange traded futures and bilateral contracts, this paper contributes to the scarce literature on this important topic and provides guidelines as to the optimal choice of models for this task. Last but not least, we conduct statistical tests for the significance of the difference in forecasting accuracy of the models and use a much longer forecast evaluation period than typically considered. Both issues have been often downplayed in the electricity price forecasting literature – but as Weron [1] argues – they both play a key role in performing a fair comparison of forecasting models. In particular, the three year test period used here (January 2011 – December 2013) allows to thoroughly evaluate the models under different market conditions and significantly reduces the risk of selecting an exceptionally favorable (or unfavorable) time period.

The remainder of the article is structured as follows. In Section II we describe the data used in this empirical study. In Sections III and IV we introduce the time series models that are calibrated either to aggregated or disaggregated data. In Section V we briefly describe the error measures and the forecasting scheme, then in Section VI we compare the out-of-sample forecasting performance of the models, with respect to the data aggregation level and inclusion of fundamental variables. Finally, in Section VII we conclude.

II. The data

The UK power market is chosen as the test ground for three reasons. Firstly, it is one of the most mature wholesale power markets in the world. Secondly, the APX power exchange (formerly UKPX) provides detailed information about the intra-day structure of the market, including prices and volumes for every half-hour. Thirdly, the large number of load periods per day – 48 compared to 24 for most other power markets – is important for the estimation of factor models. The approach used in this paper is consistent (in the statistical sense) only if the cross-sectional dimension is large. Although 48 values may still not guarantee consistency, in general, we will be better off using 48 half-hourly prices than 24 hourly prices.

A panel of 96 half-hourly electricity prices was constructed using the data downloaded from the APX web site (www.apxgroup.com) and spans the period from April 22nd, 2009 to December 31st, 2013. The panel consists of 48 volume-weighted prices and 48 spot prices. APX performs volume-weighting over three types of contracts: half-hourly, two hour block and four hour block contracts. The arithmetic average of these volume-weighted prices (48 half-hourly prices for a particular day) yields what APX calls the *baseload electricity price*, for details see https://www.apxgroup.com/market-results/apx-power-uk/ukpx-rpd-index-methodology/. We follow the same approach when considering the disaggregated models (for model definitions see Table I) – we predict the half-hourly volume-weighted prices and then take their arithmetic average as the forecast of the baseload price. Note that we do not perform the weighting ourselves, but work directly with volume-weighted prices. Note also that the 48 spot prices are only used to enrich the panel when calibrating factor models and allow for extraction of information that is not included in volume-weighted prices, but that may be relevant for predicting baseload prices. Initially we have also included 48 system buy prices and 48 system sell prices in the panel, but this did not lead to more accurate forecasts of the baseload prices and in the end we have decided to limit the panel to 96 prices.

The dataset is further expanded to include the average daily UK system demand for electricity (the arithmetic average of the 48 half-hourly values of Indicated Demand, as provided by ELEXON, see www.bmreports.com; originally reported in MW) and daily closing prices of three fundamental variables representing electricity generation costs (source: Reuters EcoWin):

- natural gas (National Balancing Point, NBP, day-ahead price index in GBp/Therm, i.e. pence per Therm),
- coal (API2 price in GBP/t; converted from USD using the Bank of England USD/GBP reference rate),
- CO_2 emissions (European Climate Exchange, ECX, Carbon Phase 3 nearest-to-delivery futures contract closing prices in GBP/t; converted from EUR using the Bank of England EUR/GBP reference rate).

As is quite common in the electricity price forecasting literature (for a review see [1]), a logarithmic transformation is applied to all five daily time series to limit the influence of price spikes and decrease the variance. The log-prices and log-demand are depicted in Figure 1. Note that, due to the relatively low prices of CO_2 emissions in 2012 and 2013, we have plotted them in a separate panel.

TABLE I
Model types and notation

Symbol	Model description
AR	The benchmark – a univariate AR model of baseload (i.e. average daily) prices P_t
ARX	The ARX model of baseload prices P_t and fundamental variables $\mathbf{X_t}$
AR_H	The 'disaggregated AR model', i.e. a set of 48 univariate AR models of half-hourly volume-weighted prices $P_{k,t}$
ARX_H	The 'disaggregated ARX model', i.e. a set of 48 univariate ARX models of half-hourly volume-weighted prices $P_{k,t}$ and fundamental variables $\mathbf{X_t}$
PC_N	The VAR model of N factors $F_{n,t}$
PC_NX	The VARX model of N factors $F_{n,t}$ and fundamental variables $\mathbf{X_t}$

Note: All computations in this study are performed on logarithms of prices and demand. Hence, the symbols P_t, $P_{k,t}$ and $\mathbf{X_t}$ refer to log-prices and log-demand. Likewise, the factors $F_{n,t}$ are obtained from the log-prices.

We use the last three years (exactly 756 business days) to evaluate the out-of-sample forecasting performance. For each day in the evaluation period, we roll the calibration window of 386 business days (which corresponds to circa 1.5 calendar years plus 5 business days for AR lags) forward by one day to ensure that all models are estimated on a sample of the same size. We have also tested other window lengths (corresponding to one and two calendar years), however, the best results were obtained for the 1.5-year window. The forecast horizons range from one to 45 business days; the latter horizon corresponds to just over two calendar months.

III. The models

In this article we focus on autoregressive (AR) models, both with and without fundamental variables. Since a stable AR(q) process has a moving average representation, it will return to its mean after any shock, even for $q > 1$. The dynamics of the return to the process mean depends on the model parameters and the lag order. For each calibration window and each model we choose the lag order, $1 \leq q \leq 5$, based on the Akaike information criterion (AIC) [4]. Note that $q = 5$ business days corresponds to a calendar week. All considered models are estimated with the Ordinary Least Squares (OLS) method.

To model the seasonal pattern of the process mean, we have initially extended the AR models to include deterministic variables: a constant and the number of daylight hours. However, after a series of extensive empirical tests we have come to the conclusion that more accurate electricity price forecasts can be obtained for models without the daylight hours variable and, hence, have not used it in the end. The rationale for this approach stems also from the fact that annual seasonality is not very apparent in UK electricity spot prices in the considered period, see Figure 1. Note also that there is no need for an additional short-term seasonal component distinguishing between the day types: working day vs. weekend vs. holiday, since – like in [10] – we are considering only business day (i.e. working day) data.

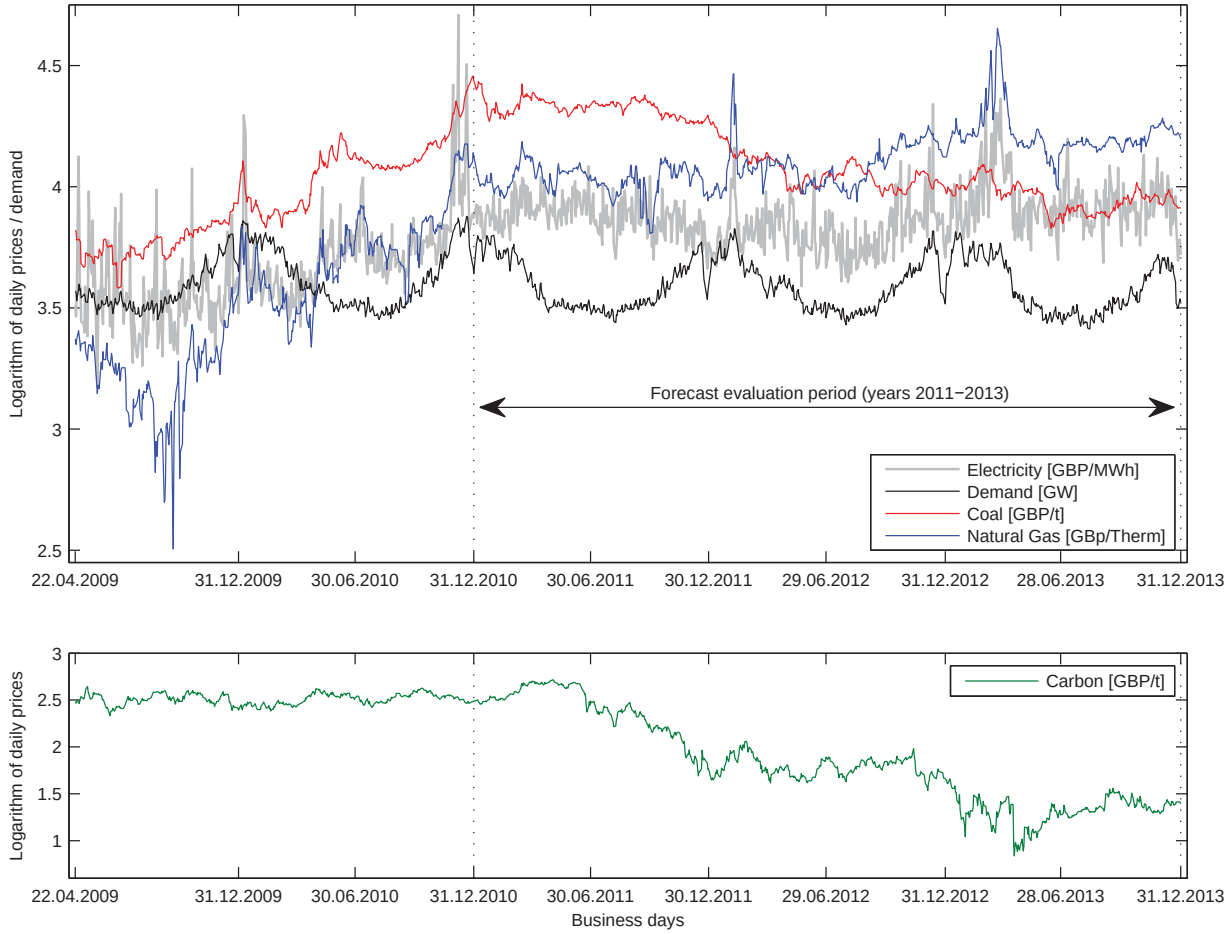

Fig. 1. *Top panel*: The logarithm of the APX baseload electricity prices (in GBP/MWh), average daily Indicated Demand (in GW), API2 coal closing prices (in GBP/t; converted from USD) and the NBP natural gas day-ahead price index (in GBp/Therm, i.e. pence per Therm) for the period from April 22nd, 2009 to December 31st, 2013. *Bottom panel*: The logarithm of the ECX Carbon Phase 3 nearest-to-delivery futures contract prices (in GBP/t; converted from EUR) in the same period. The last three years (exactly 756 business days; indicated by the vertical dotted lines in both panels) are used to evaluate the out-of-sample forecasting performance.

A. The aggregated models of baseload prices

As the benchmark we choose an autoregressive model of baseload prices. It is denoted later in the text as **AR**, see Table I. AR models are commonly used in the literature and have been shown to perform pretty well in predicting electricity spot prices, see e.g. [4], [30]–[32]. The AR model uses only the aggregated data, i.e. the baseload prices. Hence, it is suitable for comparison of aggregated and disaggregated modeling approaches.

In this model, we describe the logarithm of the baseload price, P_t, by the following AR process:

$$P_t = \alpha + \sum_{i=1}^{q} \beta_i P_{t-i} + \varepsilon_t, \tag{1}$$

where α is a constant, β_i are the autoregressive parameters and ε_t is white noise. The lag order is estimated using AIC, independently for each calibration window. The maximum lag is $q = 5$ working days, which corresponds to one calendar week. The assumption is in line with earlier works of [4] and [31], who used lags of up to one week, when forecasting California and Nord Pool spot prices. Note that in this study all lags from 1 up to q are included in the models.

We next expand the benchmark model to include exogenous variables: generation costs (prices of coal, gas and CO_2 emission rights) and Indicated Demand. In the **ARX** model we represent the logarithm of the baseload price, P_t, by the following process:

$$P_t = \alpha + \Gamma \mathbf{X_t} + \sum_{i=1}^{q} \beta_i P_{t-i} + \varepsilon_t, \tag{2}$$

where α is a constant, $\mathbf{X_t}$ is an $M \times 1$ vector of exogenous variables, Γ is a $1 \times M$ vector of parameters, β_i are the autoregressive parameters and ε_t is white noise. The number of fundamental variables, M, depends on the model specification and varies between 1 and 4.

In order to calculate forecasts of electricity prices, we first need to generate forecasts of fundamental variables. In this study, we model each of the fundamental variables, $X_{m,t}$, by

an AR(q), $1 \leq q \leq 5$, process of its own:

$$X_{m,t} = \tilde{\alpha} + \sum_{i=1}^{q} \tilde{\beta}_i X_{m,t-i} + \tilde{\varepsilon}_{m,t}, \quad m = 1, ..., 4, \quad (3)$$

where $\tilde{\alpha}$ is a constant, $\tilde{\beta}_i$ are the autoregressive parameters and $\tilde{\varepsilon}_t$ is white noise. Like in the models for electricity prices, the lag order is chosen based on AIC, separately for each calibration window and each fundamental variable.

B. The disaggregated models of half-hourly prices

In order to test if the forecasts based on disaggregated data are more accurate, we consider two 'disaggregated models' – counterparts of the univariate **AR** and **ARX** models defined in eqns. (1) and (2), respectively. The $\mathbf{AR}_H$ model is a set of 48 separate AR(q) models, one for each half-hourly load period. In this model, the half-hourly volume-weighted prices, $P_{k,t}$, are described by:

$$P_{k,t} = \alpha_k + \sum_{i=1}^{q} \beta_{k,i} P_{k,t-i} + \varepsilon_{k,t}, \quad k = 1, ..., 48, \quad (4)$$

where α_k and $\beta_{k,i}$ are the counterparts of α and β_i in eqn. (1). The baseload log-price forecasts, $\hat{P}_t$, are calculated as the average of the disaggregated, half-hourly log-price forecasts, $\hat{P}_{k,t}$:

$$\hat{P}_t = \log\left(\frac{1}{48}\sum_{k=1}^{48} \exp(\hat{P}_{k,t})\right). \quad (5)$$

Similarly, we define the $\mathbf{ARX}_H$ model of half-hourly volume-weighted prices and fundamental variables:

$$P_{k,t} = \alpha_k + \Gamma_k \mathbf{X_t} + \sum_{i=1}^{q} \beta_{k,i} P_{k,t-i} + \varepsilon_{k,t}, \quad (6)$$

where α_k, Γ_k and $\beta_{k,i}$ are the counterparts of α, Γ and β_i in eqn. (2). The choice of the vector of fundamental variables, $\mathbf{X_t}$, follows the same rules as in the case of the **ARX** model, with the number of fundamental variables varying between 1 and 4. Like for the $\mathbf{AR}_H$ model, the baseload log-price forecasts are calculated using formula (5).

C. Factor models

If we want to explore the intra-day correlations of electricity prices, we need to use dimension reduction methods. A straightforward application of a multivariate framework – like vector autoregression – would lead to a large number of parameters and could result in over-fitting, i.e. small in-sample residuals and large out-of-sample errors [8], [28]. For instance, in a VAR(q) model of half-hourly data, there will be $1 + 48q$ parameters in each equation.

Instead we suggest to use factor models, with factors estimated as principal components. Such models have been successfully applied for forecasting aggregated data [24], [29]. If we treat the half-hourly electricity spot prices as a panel then we can use the approach described in [24], [33], [34]. It was shown that the principal component estimation method is consistent for large dimensional models where both of the dimensions – time and the number of series – tend to infinity.

The main assumption of factor models is that all variables in the panel, $P_{k,t}$, co-move and depend on a small set of common factors, $F_{n,t}$. The individual series $P_{k,t}$ can be modeled as a linear function of N principal components $F_{n,t}$ and an idiosyncratic component $\nu_{k,t}$:

$$P_{k,t} = \sum_{n=1}^{N} \Lambda_{k,n} F_{n,t} + \nu_{k,t}. \quad (7)$$

The parameters $\Lambda_{k,n}$ are called *factor loadings*, as they describe the effect of the n-th factor, $F_{n,t}$, on the k-th variable in a panel, $P_{k,t}$. It was shown in [24] and [34] that $F_{n,t}$ can be consistently estimated with the eigenvectors corresponding to the N largest eigenvalues of the matrix $\mathbf{P}'\mathbf{P}$ multiplied by $\sqrt{T}$, where $t = 1, ..., T$.

In order to predict future values of half-hourly prices, we need to forecast both the common factors, $F_{n,t}$, and the idiosyncratic components, $\nu_{k,t}$ [35]. Although the factors are contemporaneously orthogonal, due to normalization assumptions, they may be still inter-temporally correlated. Hence, it seems reasonable to model them jointly. Moreover, they may depend on some other variables, such as fuel prices, CO_2 emission costs or the level of demand. At the same time, the idiosyncratic components can be only weakly correlated across periods and therefore should be modeled separately, for each half-hour. Moreover, they cannot depend on any fundamentals because all the co-movement between half-hours is captured by the factors. Once the disaggregated model is estimated, the baseload log-price forecast can be obtained by averaging the half-hourly log-price forecasts using eqn. (5).

We use an autoregressive model, AR(q), to describe and forecast the idiosyncratic component for each half-hourly load period $k = 1, ..., 48$:

$$\nu_{k,t} = \sum_{i=1}^{q} \phi_{k,i} \nu_{k,t-i} + \xi_{k,t}, \quad (8)$$

where $\phi_{k,i}$ are the autoregressive parameters and $\xi_{k,t}$ are the white noise terms. The lag order, $1 \leq q \leq 5$, is chosen based on AIC, separately for each calibration window. In model (8), neither deterministic nor fundamental variables are included.

The common factors are assumed to follow a vector autoregressive VAR(q) model, $1 \leq q \leq 5$. We distinguish two model specifications, depending on the set of variables used. In the first one – denoted by $\mathbf{PC}_N$, where N is the number of factors – only the factors are included:

$$\mathbf{F_t} = \mathbf{A} + \sum_{i=1}^{q} \mathbf{B}_i \mathbf{F_{t-i}} + \zeta_t. \quad (9)$$

Here $\mathbf{A}$ denotes an $N \times 1$ vector of deterministic coefficients, $\mathbf{B}_i$ are $N \times N$ matrices of autoregressive parameters, $\mathbf{F_t} = [F_{1,t}, ..., F_{N,t}]'$ is an $N \times 1$ vector of the factors and ζ_t is white noise.

In the second specification – denoted by $\mathbf{PC}_N\mathbf{X}$ – the fundamental variables, $\mathbf{X_t}$, are added to the model as exogenous variables:

$$\mathbf{F_t} = \mathbf{A} + \Gamma \mathbf{X_t} + \sum_{i=1}^{q} \mathbf{B}_i \mathbf{F_{t-i}} + \zeta_t, \quad (10)$$

where $\mathbf{A}$, $\mathbf{B}_i$, $\mathbf{F_t}$ and ζ_t are defined in eqn. (9), $\mathbf{X_t}$ is an $M \times 1$ vector of exogenous variables and Γ is a $1 \times M$ vector of parameters.

IV. Common factors and their interpretation

Common factors are estimated using principal component analysis (PCA). This approach is based on least squares and is aimed at minimizing the sum of squared idiosyncratic components:

$$(\hat{\mathbf{F}}, \hat{\mathbf{\Lambda}}) = \arg\min \sum_{t=1}^{T} \sum_{k=1}^{K} \left(P_{k,t} - \sum_{n=1}^{N} \Lambda_{k,n} F_{n,t} \right)^2 . \quad (11)$$

As the outcome, the estimators of common factors are the common vectors of the matrix $\mathbf{P}'\mathbf{P}$ corresponding to the N largest eigenvalues, multiplied by $\sqrt{T}$.

The optimization problem (11) does not have a unique solution and, hence, the estimators are neither locally nor globally identifiable. Therefore, some constrains need to be imposed in order to ensure the local uniqueness of the solution. The most popular restriction is contemporaneous orthogonality of common factors. Unfortunately, any non degenerated, linear transformation of principal components and appropriately transformed loadings will satisfy (11). Therefore, estimated factors do not have any direct economic interpretation.

In order to give meaning to the estimated common factors, the factors loadings need to be analyzed. The values of the factor loadings indicate which factors have strong effect on which variables. Hence, may help to understand the processes captured by the common factors.

In the electricity price forecasting literature [17], [19], [27], [29], models with two or three factors are typically used. Here, we evaluate the performance of PC models with the number of factors ranging from two to five. As we will see later in the text, the optimal number of factors changes with the forecast horizon.

The time path of the first factor, $F_{1,t}$, and the values of its loadings, $\Lambda_{k,1}$, indicate that it describes the price level. This can be seen in Figure 2, where the results of estimating the first three factors from the full dataset (i.e. from April 22nd, 2009 to December 31st, 2013) of electricity prices are displayed. The above interpretation is supported by a very high correlation coefficient, $\rho = 0.98$, between the first factor and the logarithm of the baseload price. Looking at the loadings of the second factor, we can interpret it as the spread between peak and off-peak hours. Large values of the second factor correspond to days characterized by low off-peak prices and high peak prices. Finally, the third factor seems to be responsible for the spread between the morning and evening peaks. As expected, the third component is highly seasonal, as the difference between the peak hours is much more visible during winter days than summer days. The remaining two factors are difficult to interpret.

V. Evaluating the forecasting performance

A. Evaluation metrics

In this section, we examine, whether using the intra-day information and fundamental variables improves the forecast accuracy. We consider different forecasting horizons, ranging from one to 45 business (i.e. working) days. Hourly or half-hourly day-ahead forecasts are typically used for forecast comparison in power market studies [1]–[5]. However, longer forecast horizons are also very important. For instance, mid-term forecasts from a few days up to a few months ahead are used for planning purposes (such as the adjustment of mid-term schedules and allocation of resources), risk management (including balance sheet calculations) and the valuation of exchange traded futures and bilateral contracts [6], [7], [9], [10]. In most cases, these tasks concern the baseload (or the peakload) electricity price price. In particular, the most common underlying instrument of exchange traded power derivatives is the baseload electricity price. Hence, like in [10], we consider here short- and mid-term forecasting horizons of baseload electricity prices.

The forecasting performance is measured using root mean squared errors (RMSE). For a given forecast horizon, $h = 1, ..., 45$ business days, we compute:

$$\text{RMSE}_p(h) = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{p}_{t+h|t} - p_{t+h})^2}, \quad (12)$$

where $\hat{p}_{t+h|t}$ is the baseload price forecast for day $t+h$ made on day t, p_{t+h} is the actual baseload price recorded on day $t+h$ and $T = 756$ days (i.e. the whole out-of-sample evaluation period, see Figure 1). The rationale for choosing a quadratic error metric (like RMSE) over a linear metric (like MAPE) stems from the fact that all models in this study are estimated using OLS, which minimizes the squared distance between the model and the data. However, due to the popularity of the Mean Absolute Percentage Error (MAPE) in the engineering literature, for illustrative purposes we also report MAPE values:

$$\text{MAPE}_p(h) = \frac{1}{T} \sum_{t=1}^{T} \frac{|\hat{p}_{t+h|t} - p_{t+h}|}{p_{t+h}}. \quad (13)$$

Note that both error measures concern prices, not log-prices. The baseload price forecasts are simply computed by taking the exponent of the log-price forecasts:

$$\hat{p}_{t+h|t} = \exp(\hat{P}_{t+h|t}), \quad (14)$$

which in turn are obtained as a result of using one of the models described in Section III.

The forecasts of the considered models are further evaluated on the basis of the Diebold-Mariano (DM) test [36]; for uses and abuses see [37]. The test allows to compare the forecasts of a pair of models and indicates, which statistically outperform the other. For each forecasting technique, we calculate the loss differential series $d_t = L(\varepsilon_{model1,t}) - L(\varepsilon_{model2,t})$, with the quadratic loss function $L(\varepsilon_t) = \varepsilon_t^2$. We then conduct the DM tests for significance of differences. Note that we perform one-sided DM tests, with the null hypothesis $H_0 : E(d_t) \leq 0$. Hence, when the p-value is smaller than the chosen significance level (e.g. $\alpha = 5\%$ or 10%), we can conclude that the forecasts of $model1$ are significantly better than the forecasts of $model2$.

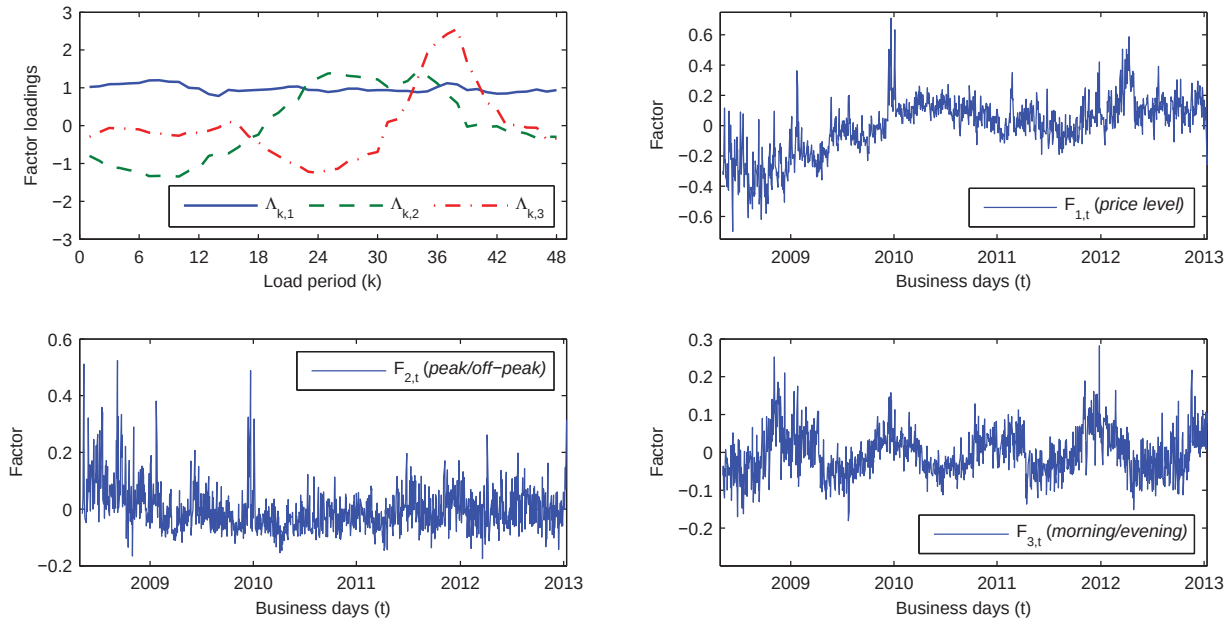

Fig. 2. Results of estimating a factor model to electricity prices from the period April 22nd, 2009 – December 31st, 2013: factor loadings $\Lambda_{k,j}$ ($j = 1, .., 3$; *top left panel*) and factors $F_{1,t}$, $F_{2,t}$ and $F_{3,t}$ (*top right and bottom panels*). The first factor may be interpreted as the price level, the second as the spread between peak and off-peak hours and the third as the spread between the morning and evening peaks.

The main assumption of the DM test is the stationarity of the loss differential series. In this study we use a rolling window of a constant length for which, in contrast to an expanding window, the parameters do not converge to their pseudo-true values (and this is one of the potential reasons for the non-stationarity of forecast errors). Moreover, the existence of the unit root of the one and 45 step-ahead forecast errors of the AR model was checked with the Augmented Dickey-Fuller (ADF) test [38]. The test rejected the null of the unit root and hence gave no reason to question covariance-stationarity of the forecast errors. It should be noted, however, that the lack of the unit root does not imply that forecast errors are not autocorrelated. The potential autocorrelation should be taken into account when estimating the variance of the loss differential. Here, we follow the approach of Diebold and Mariano [36] and use the spectrum at frequency zero of the loss differential as a robust estimator of its variance.

B. The forecasting scheme

We estimate all model parameters using information provided by the moving calibration window of a constant length (386 business days, which corresponds to 1.5 calendar years plus 5 days for AR lags). Once the parameters of models **AR**, **ARX**, $\mathbf{AR}_H$ and $\mathbf{ARX}_H$ are estimated using OLS, the h step-ahead forecasts of the baseload log-prices are computed sequentially by applying the law of iterated expectations, as in [39]. For the disaggregated models – $\mathbf{AR}_H$ and $\mathbf{ARX}_H$ – the baseload log-price forecasts are computed using eqn. (5). Note that for the ARX-type models the forecasts of the logarithms of the fundamental variables have to be computed first.

The procedure for the factor models – $\mathbf{PC}_N$ and $\mathbf{PC}_N\mathbf{X}$ – is slightly more complicated. First, for each calibration window the factors, $F_{n,t}$, and the factor loadings, $\Lambda_{k,n}$, are estimated from eqn. (7). Then, the factors are used to estimate the parameters of models (9) or (10). Once the models are estimated, factor forecasts $\hat{F}_{n,t+h|t}$ are computed sequentially, like for the non-factor models. Next, an analogous approach is applied to the idiosyncratic component. For each calibration window, the parameters of model (8) are estimated and used in sequential forecasting of future values of the idiosyncratic component, $\hat{\nu}_{k,t+h|t}$. When the forecasts of the common factors and the idiosyncratic components are available, they are used to estimate future values of the half-hourly log-prices, $P_{k,t}$, according to eqn. (7). Then the baseload log-price forecasts are computed using eqn. (5). Finally, for all six types of models, the h step-ahead forecasts of baseload log-prices, $\hat{P}_{t+h|t}$, are converted into baseload price forecasts, $\hat{p}_{t+h|t}$, using eqn. (14).

VI. Results

A. Forecasting horizons and models

The baseload price forecasts are compared for forecast horizons ranging from $h = 1$ to 45 business days. In Tables II and III the results are presented for all short-term horizons ($h = 1, ..., 5$; covering the nearest week) and three selected mid-term horizons ($h = 10, 25, 45$; equivalent to two calendar weeks, and one and two calendar months, respectively). In Tables IV and VI the forecasting horizons are aggregated into three ranges: (i) from 1 to 5 business days, (ii) from 6 to 25 business days and (iii) from 26 to 45 business days.

Due to space limitations, all four tables report on the forecasting performance of selected subsets of the full set of 54 models:

TABLE II
ROOT MEAN SQUARED ERRORS (RMSE_p) AND p-VALUES OF THE DM TEST (VS. THE BEST MODEL IN EACH COLUMN) FOR SHORT AND MID-TERM FORECASTING HORIZONS RANGING FROM 1 TO 45 BUSINESS DAYS.

$\mathbf{X_t}$	Model	Forecasting horizon (business days)								Forecasting horizon (business days)							
		1	2	3	4	5	10	25	45	1	2	3	4	5	10	25	45
		RMSE_p (for the AR model) and relative RMSE_p (vs. the AR model)								*p-values of the DM test (vs. the best model)*							
	AR	*4.270*	*4.687*	*4.842*	*4.969*	*5.084*	*5.728*	*6.528*	*6.998*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	AR_H	0.995	0.994	0.996	0.995	0.995	0.984	0.978	0.981	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00
	PC_4	0.970	0.951	0.953	0.955	0.957	0.941	0.939	0.890	0.00	0.04	0.03	0.02	0.02	0.01	0.03	0.00
Gas	ARX	0.964	0.953	0.954	0.958	0.963	0.944	0.936	0.877	0.01	0.02	0.00	0.00	0.00	0.00	**0.06**	0.03
	ARX_H	0.946	0.935	**0.937**	**0.937**	**0.940**	**0.927**	0.939	0.884	**0.38**	**0.45**	—	—	—	—	**0.06**	0.00
	PC_4X	0.977	0.960	0.962	0.957	0.957	0.941	0.929	0.889	0.01	0.01	0.00	0.01	0.02	0.01	**0.12**	0.00
Coal	ARX	0.997	0.993	0.993	0.991	0.989	0.968	0.933	0.883	0.00	0.00	0.00	0.01	0.02	0.05	0.00	**0.14**
	ARX_H	0.991	0.987	0.988	0.984	0.980	0.954	0.930	0.891	0.00	0.00	0.01	0.02	**0.06**	**0.14**	0.00	0.05
	PC_4X	1.009	1.005	1.008	0.996	0.989	0.951	**0.912**	0.883	0.00	0.00	0.00	0.00	0.02	**0.15**	—	**0.14**
CO_2	ARX	1.005	1.007	1.007	1.010	1.013	1.010	1.024	1.019	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ARX_H	0.997	0.997	0.999	0.999	1.000	0.993	1.011	1.012	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	PC_4X	1.010	1.007	1.005	0.994	0.983	0.950	0.948	0.947	0.00	0.00	0.00	0.01	0.04	**0.16**	0.01	0.00
Demand	ARX	1.013	1.025	1.034	1.039	1.048	1.066	1.060	1.009	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ARX_H	1.004	1.011	1.021	1.025	1.033	1.038	1.027	0.982	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	PC_4X	1.017	1.023	1.034	1.027	1.025	1.001	0.956	0.916	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gas, Coal	ARX	0.964	0.950	0.950	0.952	0.956	0.938	0.965	0.912	0.00	0.03	**0.15**	**0.12**	**0.08**	**0.14**	0.00	0.00
	ARX_H	**0.944**	**0.934**	0.939	0.942	0.945	0.937	0.990	0.938	—	—	**0.40**	**0.29**	**0.26**	**0.13**	0.00	0.00
	PC_4X	0.977	0.966	0.977	0.979	0.982	0.954	0.972	0.927	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Gas, Demand	ARX	0.979	0.976	0.980	0.982	0.987	0.963	0.938	**0.870**	0.00	0.00	0.00	0.00	0.00	0.00	0.04	—
	ARX_H	0.954	0.948	0.954	0.956	0.960	0.948	0.942	0.878	0.04	0.03	0.00	0.00	0.00	0.00	0.03	0.00
	PC_4X	0.991	0.984	0.993	0.989	0.993	0.973	0.929	0.876	0.00	0.00	0.00	0.00	0.00	0.00	**0.11**	**0.07**
Coal, Demand	ARX	1.004	1.007	1.013	1.012	1.012	0.981	0.936	0.875	0.00	0.00	0.00	0.00	0.00	0.01	0.00	**0.37**
	ARX_H	0.991	0.990	0.998	0.995	0.995	0.969	0.947	0.891	0.00	0.00	0.00	0.00	0.01	0.03	0.00	**0.07**
	PC_4X	1.015	1.018	1.031	1.029	1.031	0.986	0.939	0.887	0.00	0.00	0.00	0.00	0.00	0.01	0.00	**0.10**
Gas, Coal, Demand	ARX	0.973	0.969	0.978	0.982	0.983	0.961	0.988	0.926	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ARX_H	0.950	0.943	0.951	0.955	0.961	0.954	1.012	0.953	0.01	0.00	**0.06**	0.04	0.02	0.01	0.00	0.00
	PC_4X	0.982	0.972	0.988	0.987	0.988	0.965	0.989	0.938	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note 1: Root Mean Squared Errors, RMSE_p, for the benchmark AR model are presented; for other models, ratios of particular RMSE_p to the benchmark RMSE_p are displayed. Models are defined in Table I. The fundamental variables (i.e. $\mathbf{X_t}$) used for a particular model are listed in the first column. The best performing models in each column are emphasized in bold.
Note 2: The forecasts of a model are significantly worse than those of the best performing model (in each column) at significance level $\alpha = 5\%$ when the p-value is below 0.05. p-values for models not significantly worse than the best model are emphasized in bold.

- six pure electricity price models (AR, AR_H, PC_2, PC_3, PC_4 and PC_5; see Section III for model definitions),
- six models (ARX, ARX_H, PC_2X, PC_3X, PC_4X and PC_5X) for each of the eight sets of fundamental variables: (i) natural gas, (ii) coal, (iii) CO_2, (iv) demand, (v) natural gas and coal, (vi) natural gas and demand, (vii) coal and demand, and (viii) natural gas, coal and demand.

Overall we have considered many more models, in particular models with all possible sets of fundamental variables and models where the fundamental variables were predicted jointly with the prices (or factors), within vector autoregressive structures. However, due to the generally worse forecasting performance of these models we have decided to consider only the above 54 specifications in the final comparison.

B. Errors for individual models and forecasting horizons

The Root Mean Squared Errors (RMSE) for 27 representative models are presented in Table II; the remaining factor models (i.e. for $N = 2, 3$ and 5) were always outperformed by one of the presented models for a particular forecasting horizon ($h = 1, 2, 3, 4, 5, 10, 25$ or 45 business days). The first row shows the RMSE_p of the benchmark AR model. The next rows show each model's RMSE_p relative to the RMSE_p for the benchmark model. Values smaller than one indicate a better forecasting performance than the benchmark, while values greater than one indicate the opposite. This part of the table, however, does not provide information on the statistical significance of the differences. Hence, in the columns to the right we provide the p-values of the DM test vs. the best model in each column, i.e. for each of the eight forecast horizons. The forecasts of a model are significantly worse than those of the best performing model (in each column) at significance level $\alpha = 5\%$ when the p-value is below 0.05. For convenience, the p-values of models not significantly worse than the best model are emphasized in bold; the best models are indicated by a hyphen (—).

The performance of the models depends on both the set of fundamental variables used and the level of aggregation. When no fundamental variables are included, the factor model (PC_4) forecasts better than the disaggregated model (AR_H), which in turn is better than the benchmark AR model for all considered forecast horizons. The inclusion of fundamentals generally improves the forecasts, but not for all choices of the fundamental variables. In particular, the system-wide demand (also in combination with coal prices) and the CO_2 prices worsen the baseload price predictions, especially in the short-term. On the other hand, sets of fundamental variables which include natural gas prices in (almost) all cases lead to more

TABLE III
MEAN ABSOLUTE PERCENTAGE ERRORS (MAPE_p) FOR SHORT AND MID-TERM FORECASTING HORIZONS RANGING FROM 1 TO 45 BUSINESS DAYS.

$\mathbf{X_t}$	Model	Forecasting horizon (business days)							
		1	2	3	4	5	10	25	45
	AR	6.35%	6.99%	7.17%	7.37%	7.52%	8.35%	9.64%	10.47%
	AR_H	6.22%	6.82%	7.01%	7.20%	7.33%	8.10%	9.25%	10.17%
	PC_4	6.25%	6.82%	7.00%	7.21%	7.36%	8.07%	8.89%	9.00%
Gas	ARX	6.39%	6.98%	7.18%	7.36%	7.50%	8.22%	8.98%	8.89%
	$_H$-X	6.16%	6.70%	6.93%	7.07%	7.18%	7.96%	8.92%	8.93%
	PC_4X	6.25%	6.81%	7.05%	7.18%	7.32%	8.06%	8.84%	9.01%
Coal	ARX	6.35%	6.87%	7.11%	7.31%	7.39%	7.95%	8.54%	8.69%
	ARX_H	6.20%	6.71%	6.95%	7.07%	7.15%	7.75%	8.43%	8.71%
	PC_4X	6.35%	6.92%	7.19%	7.30%	7.39%	7.90%	**8.40%**	**8.64%**
CO_2	ARX	6.32%	6.88%	7.13%	7.30%	7.41%	8.23%	9.41%	10.18%
	ARX_H	6.15%	6.69%	6.94%	7.08%	7.15%	7.96%	9.17%	10.01%
	PC_4X	6.27%	6.78%	7.04%	7.11%	7.10%	**7.57%**	8.43%	9.27%
Demand	ARX	6.40%	7.16%	7.37%	7.65%	7.88%	9.18%	10.74%	11.19%
	ARX_H	6.29%	6.98%	7.23%	7.48%	7.71%	8.88%	10.24%	10.73%
	PC_4X	6.37%	7.05%	7.32%	7.49%	7.54%	8.37%	9.24%	9.62%
Gas, Coal	ARX	6.24%	6.76%	6.97%	7.07%	7.22%	7.93%	8.91%	8.67%
	ARX_H	**6.07%**	**6.60%**	**6.80%**	**6.93%**	**7.05%**	7.85%	9.13%	8.95%
	PC_4X	6.24%	6.81%	7.09%	7.20%	7.35%	8.01%	8.93%	8.80%
Gas, Demand	ARX	6.42%	7.08%	7.34%	7.48%	7.65%	8.36%	9.09%	8.93%
	ARX_H	6.21%	6.80%	7.06%	7.20%	7.35%	8.16%	9.05%	9.00%
	PC_4X	6.31%	6.91%	7.21%	7.33%	7.47%	8.24%	8.91%	9.01%
Coal, Demand	ARX	6.37%	6.99%	7.24%	7.44%	7.57%	8.17%	8.71%	8.71%
	ARX_H	6.20%	6.78%	7.04%	7.22%	7.37%	8.04%	8.78%	8.82%
	PC_4X	6.35%	7.02%	7.33%	7.48%	7.61%	8.14%	8.65%	8.71%
Gas, Coal, Demand	ARX	6.29%	6.90%	7.19%	7.31%	7.48%	8.14%	9.23%	8.90%
	ARX_H	6.10%	6.67%	6.93%	7.06%	7.23%	8.08%	9.43%	9.17%
	PC_4X	6.24%	6.85%	7.17%	7.26%	7.42%	8.10%	9.19%	9.00%

Note: Models are defined in Table I. The fundamental variables (i.e. $\mathbf{X_t}$) used for a particular model are listed in the first column. The best performing models in each column are emphasized in bold.

accurate forecasts than the benchmark.

Looking at the best performing models in each column, we can observe that in the short-term (up to 10 business days) the ARX_H model – which forecasts each of the half-hourly volume-weighted prices independently and then averages them to yield the baseload price – is a clear winner according to RMSE_p. For $h = 1$ and 2 the ARX_H model with $\mathbf{X_t} = \{\text{gas, coal}\}$ yields the best forecasts, but the ARX_H model with $\mathbf{X_t} = \{\text{gas}\}$ leads to equally good predictions. All the other models perform significantly worse (their p-values are below 5%). For $h = 3, 4, 5$ and 10 the situation reverts, now the ARX_H model with $\mathbf{X_t} = \{\text{gas}\}$ yields the best forecasts and the ARX and ARX_H models with $\mathbf{X_t} = \{\text{gas, coal}\}$ insignificantly worse predictions. For $h = 5$ and 10 also the ARX_H model with $\mathbf{X_t} = \{\text{coal}\}$ yields insignificantly worse forecasts, while for $h = 10$ also factor models – PC_4X with $\mathbf{X_t} = \{\text{coal}\}$ and – somewhat surprisingly – with $\mathbf{X_t} = \{\text{CO}_2\}$. For the two remaining mid-term horizons the situation changes a lot. For $h = 25$, the PC_4X model with $\mathbf{X_t} = \{\text{coal}\}$ is the best performer, but all the models with $\mathbf{X_t} = \{\text{gas}\}$ and the PC_4X model with $\mathbf{X_t} = \{\text{gas, demand}\}$ yield insignificantly worse forecasts. For the most distant horizon of 45 business days (or just over two calendar months), the univariate ARX model with $\mathbf{X_t} = \{\text{gas, demand}\}$ is the best performer, but the ARX and PC_4X models with $\mathbf{X_t} = \{\text{gas}\}$ or $\mathbf{X_t} = \{\text{coal, demand}\}$ yield insignificantly worse forecasts.

If we consider the linear measure, i.e. MAPE_p, instead of the squared one, the results change quantitatively but not qualitatively, see Table III. This time the ARX_H model with $\mathbf{X_t} = \{\text{gas, coal}\}$ yields the best forecasts for all short-term horizons ($h = 1, ..., 5$). For $h = 10$ business days, the PC_4X model with $\mathbf{X_t} = \{\text{CO}_2\}$ leads the pack, while for the most distant horizons ($h = 25$ and 45), the PC_4X model with $\mathbf{X_t} = \{\text{coal}\}$ is the best performer. For the best performing models, the MAPE values range from 6% to just over 7% in the short-term and from roughly 7.6% to a little over 8.6% in the mid-term. In the short-term these results are better than the ones obtained in [10] for the baseload prices in the Spanish power market, in some cases by as much as 2%. For $h = 10$, our results are better than those of the multivariate VARIMA model, but worse than of the univariate ARIMA model (in [10] the maximum horizon was $h = 15$ business days). It is worth noting that the MAPE_p errors in Table III are generally monotonically increasing with the length of the forecast horizon (independently for each model), a property to be expected. However, this is not so in Tables 1 and 2 in [10], quite likely due to the relatively short out-of-sample forecast evaluation period (4 months; compared with 36 months in our study).

C. The choice of fundamental variables

While Tables II and III report RMSE_p and MAPE_p errors for individual models, they concern only 8 out of 45 considered forecast horizons. To see a broader picture, in Tables IV and VI both the models and the forecasting horizons are aggregated. The former into model classes, the latter into three ranges:

TABLE IV
MEAN RANKS WITH RESPECT TO RMSE_p FOR MODELS GROUPED ACCORDING TO THE SET OF FUNDAMENTAL VARIABLES USED, FOR THREE RANGES OF THE FORECASTING HORIZONS.

$\mathbf{X_t}$	Forecasting horizon (business days)		
	1-5	6-25	26-45
—	6.044	5.158	7.730
Gas	**1.320**	**1.970**	7.230
Coal	27.334	**3.275**	**3.410**
CO_2	29.901	8.967	34.142
Demand	41.723	21.871	8.272
Gas, Coal	**1.516**	14.466	31.637
Gas, Demand	7.360	7.475	**1.347**
Coal, Demand	32.963	13.461	10.435
Gas, Coal, Demand	5.184	32.077	37.681

Note: The fundamental variables (i.e. $\mathbf{X_t}$) are listed in the first column. The two best performing models in each column are emphasized in bold. A mean rank of 1.000 would indicate that this model is the best performing model for all horizons in a given business day range.

TABLE V
MEAN ABSOLUTE PERCENTAGE ERRORS (MAPE) OF THE FUNDAMENTAL VARIABLES FOR FORECASTING HORIZONS $h = 1, ..., 10, 15, 25$ AND 45 BUSINESS DAYS.

h	Gas	Coal	CO_2	Demand
1	1.93%	0.91%	2.83%	1.77%
2	2.88%	1.37%	4.01%	2.50%
3	3.55%	1.72%	4.88%	2.87%
4	4.08%	2.01%	5.62%	3.09%
5	4.60%	2.23%	6.28%	3.22%
6	5.01%	2.44%	6.85%	3.51%
7	5.40%	2.63%	7.32%	3.81%
8	5.76%	2.78%	7.87%	4.00%
9	6.09%	2.91%	8.41%	4.15%
10	6.37%	3.10%	8.93%	4.31%
15	7.48%	3.85%	11.18%	5.19%
25	8.82%	5.12%	13.77%	6.49%
45	10.23%	6.68%	19.20%	9.40%

(i) from 1 to 5 business days, (ii) from 6 to 25 business days and (iii) from 26 to 45 business days. The procedure is the following. First, for each of the 45 forecast horizons all 54 models are ranked according to their RMSE_p. Then, for the best performing models within a model class (e.g. the best model with $\mathbf{X_t} = \{\text{gas, demand}\}$ in Table IV or the best PC_4-type model in Table VI) and all horizons within one of the ranges, a geometric average of the ranks is computed. A mean rank of 1.000 indicates that this model class is the best performing for all horizons in a given business day range.

In Table IV we compare the influence of the fundamental variables. Clearly the pure price models (AR, AR_H and PC_N) are outperformed by some of the models with fundamental variables. It is worth noticing that although the optimal sets of fundamental variables change with the forecasting horizons, natural gas is always a component of the best performing set of fundamentals. For $h \leq 25$, $\mathbf{X_t} = \{\text{gas}\}$ is the best choice. In the short-term it is closely followed by $\mathbf{X_t} = \{\text{gas, coal}\}$, but for $h = 6, ..., 25$ the second best choice is $\mathbf{X_t} = \{\text{coal}\}$. For the most distant forecasting horizons, $\mathbf{X_t} = \{\text{gas, demand}\}$ is the best choice and $\mathbf{X_t} = \{\text{coal}\}$ is second best again. Except for the intermediate range of $h = 6, ..., 25$, $\mathbf{X_t} = \{CO_2\}$ is a very bad predictor of the baseload electricity price, as are $\mathbf{X_t} = \{\text{coal}\}$, $\mathbf{X_t} = \{\text{demand}\}$ and $\mathbf{X_t} = \{\text{coal, demand}\}$ in the short-term. Models with such fundamental variables are ranked worse than the pure price models. Somewhat surprisingly also $\mathbf{X_t} = \{\text{gas, coal}\}$ and $\mathbf{X_t} = \{\text{gas, coal, demand}\}$ underperform for the more distant horizons.

What could be the reason for such an influence of the fundamental variables? To some extent it can be explained by the forecast errors made using the autoregressive model (3). In Table V we report the MAPE errors of the fundamental variables themselves for forecasting horizons $h = 1, ..., 10, 15, 25$ and 45 business days. Clearly the easiest to predict fundamental variable is the coal price – the MAPE errors range from 0.91% to 3.85% for $h = 15$ business days and 6.68% for $h = 45$ business days, very much like in [10]. The relatively stable evolution of the coal price favors it for the more distant horizons. Since we are not using future values of these prices, but forecast them like electricity prices, the lower the forecast errors over time the better predictors they are in the long-run.

The MAPE errors for natural gas are higher than for coal and range from 1.93% to 7.48% for $h = 15$ business days and 10.23% for $h = 45$ business days. Hence, models with natural gas as the only fundamental variable tend to perform worse for larger h. On the other hand, since natural gas is often the marginal fuel that sets the electricity price, in the shorter-term – when the forecast errors are not so large – it is the best predictor.

The MAPE errors for the system-wide demand are relatively high and only slightly lower than for the natural gas prices. The level of these errors does not explain the poor performance of models with demand as the only fundamental variable. However, if we take a look at Figure 1, we can observe that during the Winter the electricity price and the demand seem to be highly correlated, but in the Summer the electricity price generally does not react to the substantial decrease in demand. Most likely this phenomenon is responsible for such a poor performance of the models with demand as the only fundamental variable. Unfortunately we are not able to explain why demand combined with natural gas performs so well, see Table IV.

Finally, the MAPE errors for CO_2 are the highest of all and range from 2.83% to 11.18% for $h = 15$ business days and as much as 19.20% for $h = 45$ business days. They are much higher than the MAPE errors reported in [10] for EUA and CER certificates. Most likely this is due to the time period analyzed. Between March 2009 and April 2011 the CO_2 prices were relatively stable, see Figure 1. However, starting from June 2011 they declined rapidly until mid-2013 and then again remained stable until the end of the studied period (December 2013).

D. Aggregated vs. disaggregated models

Let us now examine whether models calibrated to disaggregated data perform better than the models calibrated to aggregated (i.e. baseload) prices. In Table VI we compare the performance of six classes of models with and without fundamental variables: AR/ARX, AR_H/ARX_H, PC_2/PC_2X, PC_3/PC_3X, PC_4/PC_4X and PC_5/PC_5X. In the short-term the disaggregated AR_H/ARX_H model is the unanimous winner with a mean rank of 1.000! Recall that such a mean rank indi-

TABLE VI
MEAN RANKS WITH RESPECT TO RMSE_p FOR MODELS GROUPED ACCORDING TO THE MODEL TYPE, FOR THREE RANGES OF THE FORECASTING HORIZONS.

Model	Forecasting horizon (business days)		
	1-5	6-25	26-45
AR / ARX	4.547	11.602	2.062
AR_H / ARX_H	**1.000**	3.820	7.842
PC_5 / PC_5X	7.955	2.910	3.399
PC_4 / PC_4X	7.721	**2.039**	2.290
PC_3 / PC_3X	7.360	2.827	**2.024**
PC_2 / PC_2X	5.553	9.102	10.454

Note: The best performing model in each column is emphasized in bold. A mean rank of 1.000 indicates that this model class is the best performing for all horizons in a given business day range.

cates that this is the best performing model class for all short-term horizons. Apparently it is advantageous to independently predict all 48 half-hourly price series and then aggregate them to yield the baseload price forecast for $h = 1, ..., 5$. However, as we increase the forecasting horizon the AR_H/ARX_H model performs worse (with respect to the other models). In the intermediate range ($h = 6, ..., 25$) the factor models with 3 or more factors dominate the ranking. The PC_4-type model yields the most accurate forecasts, but is closely followed by PC_3-type and PC_5-type; the factor model with only two factors underperforms for all h. For the most distant horizons ($h = 26, ..., 45$), the factor models with 3 or more factors still perform very well. Now PC_3-type is the best, but the univariate AR/ARX model performs nearly as well, which confirms that reducing model complexity improves the mid-term forecasting performance.

VII. CONCLUSIONS

This article examines whether using intra-day data and fundamental variables can improve forecasts of baseload electricity prices. In order to overcome the dimension problem, we use either a set of univariate AR models (one for each load period) or factor models that summarize the information contained in the panel of intra-day prices. We conduct an empirical study, which allows to assess the forecasting performance of six types of models: (i) **AR** – the benchmark AR model of baseload prices, (ii) **ARX** – the ARX model of baseload prices and fundamental variables, (iii) $\mathbf{AR}_H$ – a set of 48 univariate AR models of half-hourly volume-weighted prices, (iv) $\mathbf{ARX}_H$ – a set of 48 univariate ARX models of half-hourly volume-weighted prices and fundamental variables, (v) $\mathbf{PC}_N$ – the VAR model of $N = 2, ..., 5$ factors, and (vi) $\mathbf{PC}_N\mathbf{X}$ – the VARX model of $N = 2, ..., 5$ factors and fundamental variables. The models are compared in terms of RMSE and MAPE errors.

We should note here that the choice of the time series models used here is not exhaustive, in particular nonlinear models are not considered. However, there is mixed evidence on the forecasting performance of nonlinear models in general [40] and in electric load [41] and price forecasting in particular [1]. In fact, as Hong et al. [41] emphasize, all of the four winning teams in the load forecasting track of the Global Energy Forecasting Competition 2012 (GEFCom2012) used regression analysis to produce the winning entries. That said, there is no reason to believe that nonlinear models would yield more competitive benchmarks than the ones already used.

For the pure price models, we find that the factor model (PC_4) forecasts better than the disaggregated model (AR_H), which in turn is better than the benchmark AR model for all considered forecast horizons. The inclusion of fundamentals generally improves the forecasts, but not for all choices of the fundamental variables. In particular, the system-wide demand (also in combination with coal prices) and the CO_2 prices worsen the baseload price predictions, especially in the short-term. On the other hand, sets of fundamental variables which include natural gas prices in (almost) all cases lead to more accurate forecasts than the benchmark.

The poor performance of models with CO_2 as the fundamental variable can be explained by the errors made when forecasting the carbon prices themselves. The MAPE errors for CO_2 are the highest of all the considered fundamental variables, much higher than the MAPE errors reported in [10] for EUA and CER certificates. Most likely this is due to the evolution of CO_2 prices in the analyzed time period. Between March 2009 and April 2011 the CO_2 prices were relatively stable, but starting from June 2011 they declined rapidly until mid-2013 and then again remained stable until the end of the studied period (December 2013). Obviously this structural change cannot be adequately predicted by a simple AR(q) model.

The somewhat surprising, poor predictive performance of the system-wide demand cannot be explained by the level of the forecast errors. They are relatively high but still lower than for the natural gas prices (which is a very good predictor for all forecasting horizons). However, analyzing the time evolution of the electricity price and demand time series, we can observe that during the Winter the electricity price and the demand seem to be highly correlated, but in the Summer the electricity price generally does not react to the substantial decrease in demand. Most likely this phenomenon is responsible for such a poor performance of the models with demand as the only fundamental variable. Unfortunately we are not able to explain why demand combined with natural gas performs so well for the most distant forecasting horizons.

Now let us comment on the model structure. In the class of models with fundamental variables, we find that in the short-term (up to a few business days ahead) the disaggregated model which independently predicts the half-hourly volume-weighted prices and then takes their average to yield baseload price forecasts (i.e. AR_H/ARX_H) is the best performer. However, in the mid-term, factor models which extract information from the panel of intra-day prices – especially PC_3-type and PC_4-type – lead to significantly (as measured by the Diebold-Mariano test) better baseload price forecasts. Interestingly, for $h = 6, ..., 25$ business days, the PC_4-type model is the best performer, but for $h = 26, ..., 45$ business days, the PC_3-type model takes the lead, with the simple univariate AR/ARX model following closely by.

Summing up, there is clear evidence that using intra-day prices improves the short- and mid-term forecasts of baseload electricity prices in the UK market. However, the optimal

model structure is not the same across the forecasting horizons – the more distant the forecasting horizon, the simpler should the model structure be. The results for including fundamental variables are less straightforward. On one hand, some fundamental variables – especially natural gas prices (in the short-term) and coal prices (in the mid-term) – provide significant gains. On the other, the remaining variables – especially the CO_2 prices – do not improve the price forecasts at all, at least in the time period considered in this study (April 2009 – December 2013).

Overall, this paper contributes to the scarce literature on the important topic of mid-term electricity price forecasting and provides guidelines as to the optimal choice of models for this task. Given that mid-term forecasts play a crucial role in planning activities (such as the adjustment of mid-term schedules and allocation of resources), risk management and the valuation of exchange traded futures and bilateral contracts, the discussed results and insights may be useful not only for academics, but also for practitioners managing portfolios of electricity contracts.

REFERENCES

[1] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, pp. 1030–1081, 2014.

[2] S. Chan, K. Tsui, H. Wu, Y. Hou, Y.-C. Wu, and F. Wu, "Load/price forecasting and managing demand response for smart grids," *IEEE Signal Processing Magazine*, pp. 68–85, September 2012.

[3] T. Hong, "Energy forecasting: Past, present, and future," *Foresight*, pp. 43–48, Winter 2014.

[4] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. John Wiley & Sons, Chichester, 2006.

[5] H. Zareipour, *Price-based energy management in competitive electricity markets*. VDM Verlag Dr. Müller, 2008.

[6] M. Burger, B. Graeber, and G. Schindlmayr, *Managing energy risk: An integrated view on power and other energy markets*. Wiley, 2007.

[7] R. Carmona and M. Coulon, "A survey of commodity markets and structural models for electricity prices," in *Quantitative Energy Finance: Modeling, Pricing, and Hedging in Energy and Commodity Markets*, F. Benth, V. Kholodnyi, and P. Laurence, Eds. Springer, 2014, pp. 41–83.

[8] C. Garcia-Martos and A. Conejo, "Price forecasting techniques in power systems," in *Wiley Encyclopedia of Electrical and Electronics Engineering*. Wiley, 2013, pp. 1–23, (DOI: 10.1002/047134608X.W8188).

[9] X. Yan and N. A. Chowdhury, *Electricity market clearing price forecasting in a deregulated market: A neural network approach*. VDM Verlag Dr. Müller, 2010.

[10] C. Garcia-Martos, J. Rodriguez, and M. Sanchez, "Modelling and forecasting fossil fuels, CO2 and electricity prices and their volatilities," *Applied Energy*, vol. 101, pp. 363–375, 2013.

[11] C. Bernhardt, C. Klüppelberg, and T. Meyer-Brandis, "Estimating high quantiles for electricity prices by stable linear models," *Journal of Energy Markets*, vol. 1, pp. 3–19, 2008.

[12] K. Chan and P. Gray, "Using extreme value theory to measure value-at-risk for daily electricity spot prices," *International Journal of Forecasting*, vol. 22, pp. 283–300, 2006.

[13] S. Koopman, M. Ooms, and A. Carnero, "Periodic seasonal reg-arfima-garch models for daily electricity spot prices," *Journal of the American Statistical Association*, vol. 102, pp. 16–27, 2007.

[14] S. Schlueter, "A long-term/short-term model for daily electricity prices with dynamic volatility," *Energy Economics*, vol. 32, pp. 1074–1081, 2010.

[15] F. A. Wolak, "Market design and price behavior in restructured electricity markets: An international comparison," in *Deregulation and Interdependence in the Asia-Pacific Region, NBER-EASE, Vol. 8*, T. Ito and A. Krueger, Eds. University of Chicago Press, 2000, pp. 79–137.

[16] J. Chen, S.-J. Deng, and X. Huo, "Electricity price curve modeling and forecasting by manifold learning," *IEEE Transactions on Power Systems*, vol. 23, pp. 877–888, 2008.

[17] W. Härdle and S. Trück, "The dynamics of hourly electricity prices," SFB 649 Discussion Paper 2010-013, 2010.

[18] A. Alonso, C. Garcia-Martos, J. Rodriguez, and M. Sanchez, "Seasonal dynamic factor analysis and bootstrap inference: Application to electricity market forecasting," *Technometrics*, vol. 53, pp. 137–151, 2011.

[19] C. Garcia-Martos, J. Rodriguez, and M. Sanchez, "Forecasting electricity prices by extracting dynamic common factors: Application to the Iberian Market," *IET Generation, Transmission & Distribution*, vol. 6, no. 1, pp. 11–20, 2012.

[20] H. C. Wu, S. C. Chan, K. M. Tsui, and Y. Hou, "A new recursive dynamic factor analysis for point and interval forecast of electricity price," *IEEE Transactions on Power Systems*, vol. 28, pp. 2352–2365, 2013.

[21] D. F. Hendry and K. Hubrich, "Forecasting economic aggregates by disaggregates," European Central Bank, Working Paper Series No. 589, 2006.

[22] C. Bermingham and A. D'Agostino, "Understanding and forecasting aggregate and disaggregate price dynamics," *Empirical Economics*, vol. 46, pp. 765–788, 2014.

[23] N. Perevalov and P. Maier, "On the advantages of disaggregated data: Insights form forecasting the U.S. economy in a data-rich environment," Bank of Canada, Working Paper 2010-10, 2010.

[24] J. H. Stock and M. W. Watson, "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1167–1179, 2002.

[25] H. Lütkepohl, "Forecasting nonlinear aggregates and aggregates with time-varying weights," *Jahrbücher für Nationalökonomie und Statistik*, vol. 231, pp. 107–133, 2011.

[26] D. F. Hendry and K. Hubrich, "Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate," *Journal of Business and Economic Statistics*, vol. 29, pp. 216–227, 2011.

[27] D. Liebl, "Modeling and forecasting electricity spot prices: A functional data perspective," *The Annals of Applied Statistics*, vol. 7, no. 3, pp. 1562–1592, 2013.

[28] K. Maciejowska and R. Weron, "Forecasting of daily electricity spot prices by incorporating intra-day relationships: Evidence form the UK power market," IEEE Conference Proceedings – EEM13, art. no. 6607314, 2013.

[29] E. Raviv, K. E. Bouwman, and D. van Dijk, "Forecasting day-ahead electricity prices: Utilizing hourly prices," Tinbergen Institute Discussion Paper 13-068/III, DOI 10.2139/ssrn.2266312, 2013.

[30] A. J. Conejo, J. Contreras, R. Espínola, and M. A. Plazas, "Forecasting electricity prices for a day-ahead pool-based electric energy market," *International Journal of Forecasting*, vol. 21, pp. 435–462, 2005.

[31] R. Weron and A. Misiorek, "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models," *International Journal of Forecasting*, vol. 24, pp. 744–763, 2008.

[32] A. Misiorek, S. Trück, and R. Weron, "Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models," *Studies in Nonlinear Dynamics & Econometrics*, vol. 10, no. 3, Article 2, 2006.

[33] J. Bai and S. Ng, "Determining the number of factors in approximate factor models," *Econometrica*, vol. 70, no. 1, pp. 191–221, 2002.

[34] J. Bai, "Inferential theory for factor models of large dimensions," *Econometrica*, vol. 71, no. 1, pp. 135–171, 2003.

[35] J. Boivin and S. Ng, "Understanding and comparing factor-based forecasts," *International Journal of Central Banking*, vol. 1, 2005.

[36] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, vol. 13, pp. 253–263, 1995.

[37] F. X. Diebold, "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests," Paper prepared for Journal of Business and Economic Statistics Invited Lecture, ASSA Meeting, Philadelphia, 2014.

[38] W. H. Greene, *Econometric Analysis (5th ed.)*. Prentice Hall, New Jersey, 2002.

[39] M. Clements and D. Hendry, *Forecasting economic time series*. Cambridge University Press, 1998.

[40] M. Bessec and O. Bouabdallah, "What causes the forecasting failure of Markov-switching models? A Monte Carlo study," *Studies in Nonlinear Dynamics and Econometrics*, vol. 9, 2005, Article 6.

[41] T. Hong, P. Pinson, and S. Fan, "Global Energy Forecasting Competition 2012," *International Journal of Forecasting*, vol. 30, pp. 357–363, 2014.

Katarzyna Maciejowska received a Ph.D. in Economics from the European University Institute, Florence, Italy. Since 2011 she has been working at the Wrocław University of Technology, Poland, where she teaches both undergraduate and graduate courses in economics and econometrics. Her research focuses on econometric theory and applications, with particular emphasis on forecasting in energy markets. Her additional interests include agent-based modeling of innovation diffusion and macroeconometrics. In 2013 she has been appointed a CERGE-EI Teaching Fellow, Prague, Czech Republic.

Rafał Weron is Professor of Economics at the Wrocław University of Technology, Poland. His research focuses on developing risk management and forecasting tools for the energy industry and computational statistics as applied to finance and insurance. He is the author of the widely acclaimed *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach* (Wiley, 2006) and co-author of four other books. He has also published over 80 peer-reviewed book chapters and journal articles (most notably in top-tier *Energy Economics*, *Energy Policy*, *IEEE Transactions on Power Systems* and *International Journal of Forecasting*), including the 52-page long invited review paper on electricity price forecasting [1]. With a Ph.D. in Financial Mathematics and a habilitation ('higher doctorate') in Economics, he is periodically engaged as a consultant to financial, energy and software engineering companies.

HSC Research Report Series 2015

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *Probabilistic load forecasting via Quantile Regression Averaging on sister forecasts* by Bidong Liu, Jakub Nowotarski, Tao Hong and Rafał Weron

02 *Sister models for load forecast combination* by Bidong Liu, Jiali Liu and Tao Hong

03 *Convenience yields and risk premiums in the EU-ETS - Evidence from the Kyoto commitment period* by Stefan Trück and Rafał Weron

04 *Short- and mid-term forecasting of baseload electricity prices in the UK: The impact of intra-day price relationships and market fundamentals* by Katarzyna Maciejowska and Rafał Weron