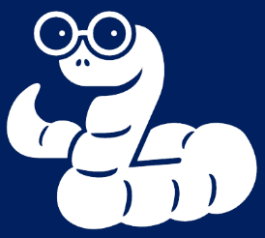

WORKing papers in Management Science

WORMS/21/04

Importance of the long-term seasonal component in day-ahead electricity price forecasting revisited: Parameter-rich models estimated via the LASSO

Arkadiusz Jędrzejewski[1]
Grzegorz Marcjasz[1]
Rafał Weron[1]

[1] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Poland

WORMS is a joint initiative of the Management Science departments of the Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

MDPI

Article

Importance of the long-term seasonal component in day-ahead electricity price forecasting revisited: Parameter-rich models estimated via the LASSO

Arkadiusz Jędrzejewski[1], **Grzegorz Marcjasz**[1], **Rafał Weron**[1]

[1] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

* Correspondence: rafal.weron@pwr.edu.pl

Version March 28, 2021 submitted to Energies

Abstract: Recent studies suggest that decomposing a series of electricity spot prices into a trend-seasonal and a stochastic component, modeling them independently and then combining their forecasts, can yield more accurate predictions than an approach in which the same parsimonious regression or neural network-based model is calibrated to the prices themselves. Here, we show that significant accuracy gains can be achieved also in the case of parameter-rich models estimated via the *least absolute shrinkage and selection operator* (LASSO). Moreover, we provide insights as to the order of applying seasonal decomposition and variance stabilizing transformations before model calibration, and propose two well-performing forecast averaging schemes based on different approaches to modeling the long-term seasonal component.

Keywords: electricity price forecasting; day-ahead market; LASSO; long-term seasonal component; variance stabilizing transformation; forecast averaging

1. Introduction

The trend-seasonal pattern of electricity spot prices, also known as the *long-term seasonal component* (LTSC), has always attracted the attention of energy analysts [1–7]. Particularly, when modeling average daily prices in the medium- or the long-term. On the other hand, the short-term *electricity price forecasting* (EPF) literature has generally ignored it and considered models with only intra-day and intra-week periodicities, as the LTSC was believed to add unnecessary complexity. Only recently, Nowotarski and Weron [8] have introduced the *seasonal component* (SC) approach and the *seasonal component autoregressive* (SCAR) models that decompose the electricity spot price series into a trend-seasonal and a stochastic component, predict them independently and combine their day-ahead forecasts. The seasonal component approach works well for autoregressive (AR) [7,9] as well as non-linear autoregressive (NARX) neural network-type models [10], in the context of point and probabilistic predictions [11]. However, the studies published to date may be criticized for utilizing only parsimonious structures with a relatively small number of explanatory variables or features. And these are known to underperform, when compared to parameter-rich models with hundreds of regressors estimated via the *least absolute shrinkage and selection operator* (LASSO) [12–16]. To our best knowledge, only two studies have treated the LTSC in the context of LASSO-estimated models [17,18]. However, no comparisons have been made between different variants of the LTSC or with analogous models that do not utilize seasonal decomposition. An open question remains whether the SC approach is also beneficial in the case of parameter-rich LASSO-estimated models and to what extent.

To this end, we perform an extensive empirical study which involves:

- 6-year-long electricity price and fundamental variables time series from two distinct power markets – Nord Pool and PJM Interconnection, providing two 3-year-long test periods of hourly-resolution.
- Two commonly used approaches to modeling the LTSC of electricity price series – one based on wavelet smoothing [2–5,7] and one on the Hodrick-Prescott (HP) filter [4,10,19,20].
- A parameter-rich, LASSO-estimated autoregressive model with nearly 130 regressors, after [21] called the LEAR model.
- The area hyperbolic sine (asinh) *variance stabilizing transformation* (VST) [22,23], which has been found to perform well when negative or close to zero electricity prices are analyzed [15,24–26].
- Two methods of combining point forecasts – one that selects the best combination of a pool of models (dubbed Best Combination, BC) and one inspired by Bayesian Model Averaging (BMA) [27]. The latter weighs combinations by the *inverse (of the) root mean squared error* (iRMSE; similarly as in [28–30]), instead of the posterior probabilities originally proposed in [27]. For notational convenience, we refer to it as BMA.
- Model validation in terms of the robust *relative mean absolute errors* (rMAE) and *relative root mean squared errors* (rRMSE) [21,31], and the Giacomini and White [32] test for significant differences in *conditional predictive ability* (CPA).

Since we utilize both seasonal decomposition (via the HP or the wavelet filter) and variance stabilization, a question arises about the order in which they should be applied. In some studies, the VST comes first [2,4,8–11], whereas in other seasonal decomposition [18]. Not having found clear recommendations in the literature, we compare both approaches.

The remainder of the paper is structured as follows. In Section 2 we briefly present the datasets. Then, in Section 3, we describe the methodology: the forecasting framework, the asinh transformation, seasonal decomposition, a parsimonious autoregressive model used as a benchmark, the LEAR model, and the two combination schemes. In Section 4 we measure forecast accuracy in terms of rMAE and rRMSE, evaluate conditional predictive ability and comment on computational complexity of the proposed methods. Finally, in Section 5, we wrap up the results and conclude.

2. Datasets

We evaluate the considered models using datasets from two major power markets. The first one is Nord Pool, a renewable energy sources dominated market in the Northern Europe, exhibiting long-term, weather-dependent fluctuations in price levels. The second is PJM Interconnection, the world's largest competitive wholesale electricity market covering Northeastern United States, with a coal-gas-nuclear generation mix.

The Nord Pool (NP) dataset, depicted in Fig. 1, comprises three time series at hourly resolution: day-ahead market system prices in EUR/MWh, day-ahead system load forecasts (called *consumption prognosis*) for four Nordic countries (Denmark, Finland, Norway and Sweden), and day-ahead wind power generation forecasts for Denmark. The data is freely available on the Nord Pool website www.nordpoolspot.com. The PJM dataset, depicted in Fig. 2, also comprises three time series at hourly resolution: day-ahead market prices in the Commonwealth Edison (COMED; located in the state of Illinois) zone in USD/MWh and two day-ahead load forecasts series – the system load and the COMED zonal load. The data is freely available on the PJM website www.pjm.com. Both datasets span the same 6-year-long time period – from 1 January 2013 to 24 December 2018 (exactly $2184 = 6 \cdot 364$ days or $52\,416 = 6 \cdot 364 \cdot 24$ hours). Note, that while the NP price is more volatile in the 3-year-long evaluation window than in the model selection window, the PJM price exhibits the opposite behavior. This will allow us to compare the models under different market conditions. Also note, that the same datasets were used in [21].

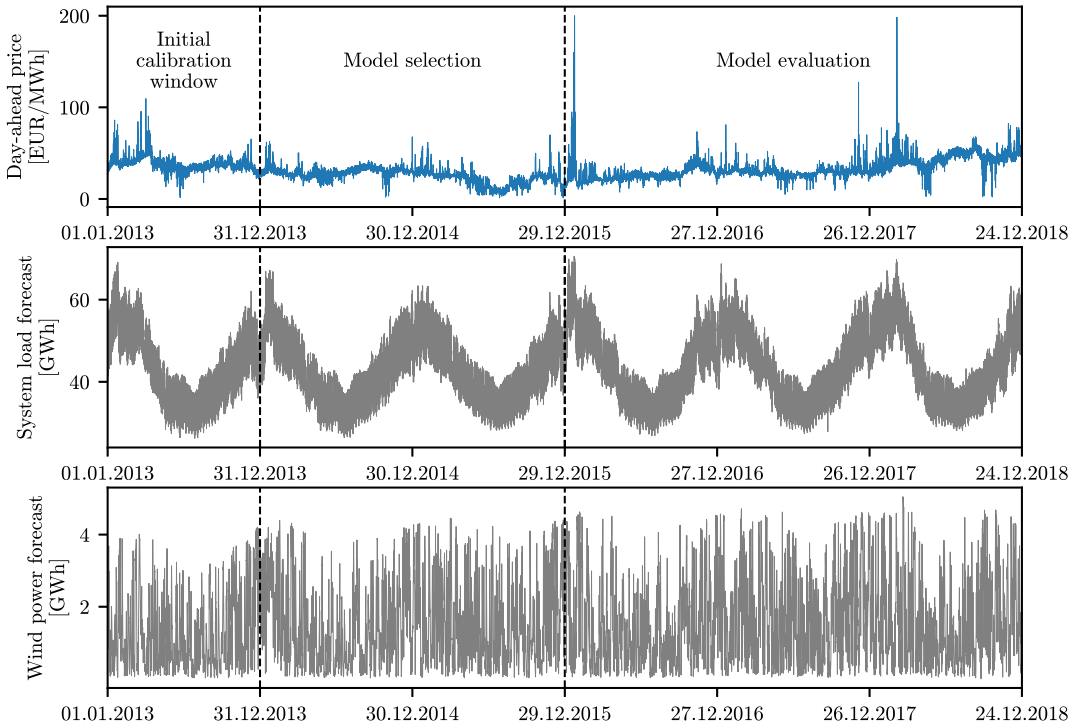

Figure 1. Day-ahead system prices in the Nord Pool electricity market (NP; *top*) and two forward looking series: day-ahead system load forecasts (*middle*) and day-ahead wind power generation forecasts for Denmark (*bottom*). The vertical dashed lines indicate the beginnings of the model selection (31.12.2013) and model evaluation (29.12.2015) windows.

3. Methodology

We implement the multivariate modeling framework, as defined in [15], in which the predictions are performed separately for each hour. To represent our variables, we use 'day $\times$ hour' matrix-like structures. The day-ahead forecasts $\hat{p}_{d,h|d-1} \equiv \hat{p}_{d,h}$ of the electricity price $p_{d,h}$ for day d and hour h are computed on day $d-1$ using all the information known up to that point in time. We utilize a rolling window scheme, i.e., our models are calibrated to data in a 364-day-long window, and once the 24 forecasts for all hours of the next day are made, the window is rolled forward 24 h. Then, the models are calibrated again, and forecasts are computed for the next 24 h. This procedure continues until the predictions for the last day in the sample are made.

Our models involve day-ahead electricity prices and two exogenous variables, see Sec. 2 and Figs. 1-2. All three time series undergo a variance stabilizing transformation (VST; see Sec. 3.1 below) and seasonal decomposition (see Sec. 3.2 below) prior to model calibration; we consider two orders of applying these transformations – VST first then seasonal decomposition or vice versa. The only exceptions are the benchmark models for which seasonal decomposition is not performed. Although we use a multivariate modelling framework, both transformations are carried out in the calibration window for the original time series at hourly resolution, just like in [8]. Moreover, since the forecasts of the exogenous variables are known one day in advance, in their case, seasonal decomposition and variance stabilization are applied to a window that is one day longer, i.e., which covers 365 days and includes the day for which the price forecasts are made.

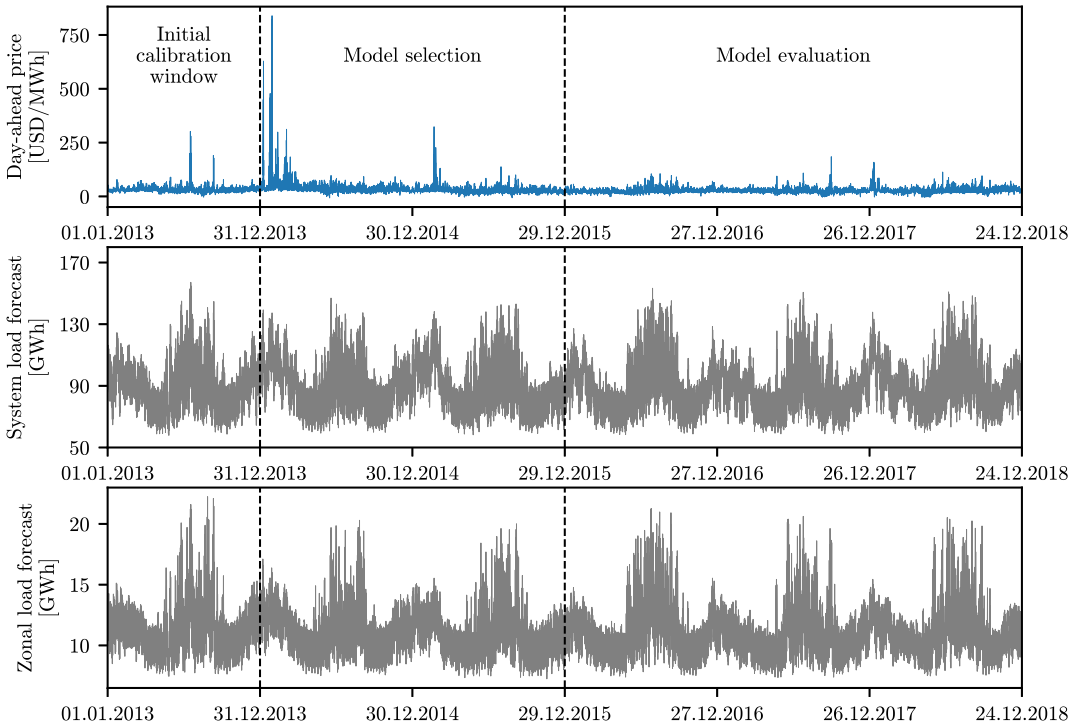

Figure 2. Day-ahead market prices in the Commonwealth Edison zone (COMED; *top*) and two forward looking series: the day-ahead system (*middle*) and COMED zone (*bottom*) load forecasts. The vertical dashed lines indicate the beginnings of the model selection (31.12.2013) and model evaluation (29.12.2015) windows.

When it comes to forecasting, like in [8,10], the LTSC of the price time series is assumed to persist into the future (see Sec. 3.2 below), whereas the stochastic component is predicted based on two models with an autoregressive structure. The first one is a parsimonious *ordinary least squares* (OLS) estimated model, built on some prior knowledge of experts (see Sec. 3.3 below), whereas the second one is a parameter-rich LASSO-estimated model (see Sec. 3.4 below).

Finally, in Section 3.5, we propose two methods of averaging/selecting forecasts. We construct our combined predictions based on performance in a 2-year-long model selection (or validation) window and compare them – like all considered models – in the 3-year-long evaluation (or test) window, see Figs. 1 and 2.

3.1. Data Transformation

A number of transformations can be used to reduce the variation in data and allow to handle close to zero or negative electricity prices. Here, following [15,18,22], we resort to the *area hyperbolic sine* (asinh), which is a simple but well performing variance stabilizing transformation (VST). Before applying it, we normalize each series $x_{d,h}$, i.e., prices or load/wind forecasts, using its median Med_τ and median absolute deviation MAD_τ in the calibration window τ:

$$y_{d,h} = \frac{1}{z_{0.75}} \frac{x_{d,h} - \text{Med}_\tau}{\text{MAD}_\tau}, \tag{1}$$

where $z_{0.75}$ is the 75th percentile of the standard normal distribution and ensures asymptotically normal consistency to the standard deviation. Such a normalization is more robust to outliers than the one based on the mean and the standard deviation, and thus it is preferred for spiky data [23]. Having normalized our variables, we apply the VST:

$$z_{d,h} = \operatorname{asinh}(y_{d,h}) = \log\left(y_{d,h} + \sqrt{y_{d,h}^2 + 1}\right). \tag{2}$$

The forecasts are obtained for models calibrated to the VST-transformed series, after applying the inverse transformation:

$$x_{d,h} = \frac{\mathrm{MAD}_\tau}{z_{0.75}} \sinh\left(z_{d,h}\right) + \mathrm{Med}_\tau, \tag{3}$$

where sinh is the hyperbolic sine. Since the latter is a nonlinear function, given random variable X, $\mathbb{E}\sinh(X)$ does not have to equal $\sinh(\mathbb{E}X)$. Hence, from the probabilistic point of view, Eq. (3) is not the correct inverse transformation. Although this problem is generally ignored in the literature, Narajewski and Ziel [33] argue that using the correct transformation yields better forecasts. On the other hand, when the forecasts are averaged, the differences between the two approaches seem to vanish [34]. Since we eventually consider forecast combinations, to keep it simple, we use Eq. (3) as the inverse transformation.

3.2. Seasonal Decomposition

Seasonal decomposition is a term that generally refers to representing a signal as a sum or a product of a periodic component and the remaining variability (also known as the stochastic component). Recent studies have shown that seasonal decomposition of day-ahead prices and separate treatment of the seasonal and stochastic components can result in more accurate forecasts generated by OLS-estimated [10,11] as well as LASSO-estimated [18] models. The approach proposed by Nowotarski and Weron [8] relies on an additive decomposition that splits the original time series, Y_t, into the *long-term trend-seasonal component* (LTSC), T_t, and the remaining stochastic component with short-term periodicity, X_t:

$$Y_t = T_t + X_t. \tag{4}$$

In our study, depending on the order of applying seasonal decomposition and the variance stabilizing transformation, Y_t may denote asinh-transformed or raw data. In the latter case, the VST is applied only to X_t. Both components are modeled independently and their predictions are combined to yield price forecasts. Like in [8–11], we consider persistent forecasts, i.e., assume that the last 24 hourly observations of the LTSC will repeat the next day:

$$\widehat{T}_{d^*+1,h} = T_{d^*,h}, \tag{5}$$

where d^* is the last day in the calibration window and the single and double time indexes satisfy $t = 24d + h$.

We consider two well-performing methods of extracting and modeling the LTSC – based on wavelet smoothing [2–5,7] and one on the Hodrick-Prescott (HP) filter [4,10,19,20]. In wavelet smoothing [1,35], the original time series is decomposed using the discrete wavelet transform into a sum of an approximation series capturing the general trend, S_k, and a number of detail series, D_k, representing higher-frequency components: $S_k + D_k + D_{k-1} + ... + D_1$, where k is the smoothing level. The LTSC is then approximated by S_k. We consider the Daubechies family of wavelets, which is frequently used in EPF studies [8–11,36–39]. More precisely, we utilize Daubechies wavelets of order 4 (denoted by 'db4'), and the smoothing level k ranges from 6 to 14. See the left panels in Fig. 3 for an illustration.

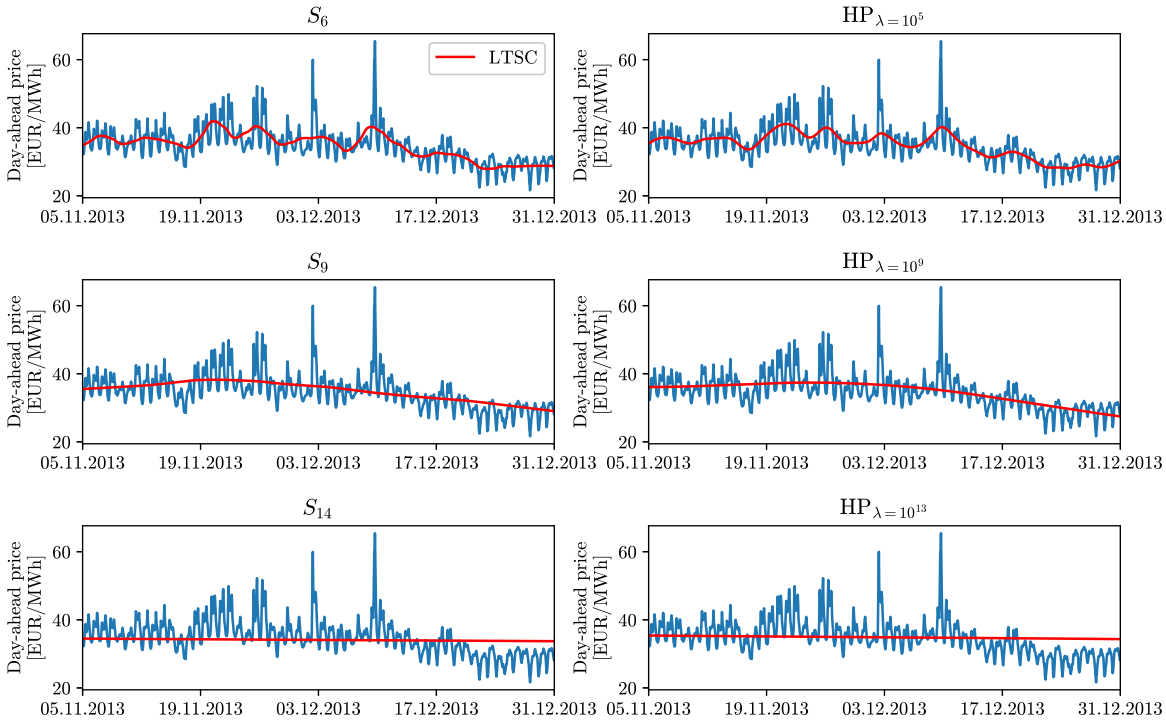

Figure 3. Illustration of the long-term trend-seasonal component (LTSC) based on wavelets S_6, S_9 and S_{14} (left column) and HP filters with $\lambda = 10^5, 10^9$, and 10^{13} (right column) fitted to Nord Pool prices in the initial calibration window, see Fig. 1. Only the last 8 weeks are displayed.

While wavelet smoothing removes layers of details, D_1, D_2, etc., the Hodrick-Prescott [40] filter simply returns T_t which minimizes squared deviations from Y_t (first term) and squared fluctuations of the smoothed series itself (second term):

$$\min_{T_t} \left\{ \sum_{t=t_0}^{\tau} (Y_t - T_t)^2 + \lambda \sum_{t=t_0+1}^{\tau-1} \Big[(T_{t+1} - T_t) - (T_t - T_{t-1}) \Big]^2 \right\}, \tag{6}$$

where t_0 and τ are respectively the beginning and the end of the calibration window in the single time index notation, and λ is the smoothing parameter. Here, we consider nine values of the latter: $\lambda = 10^k$ with $k = 5, 6, ..., 13$. The larger the λ, the smoother the resulting series is, see the right panels in Fig. 3.

3.3. The OLS-estimated Benchmark

As a benchmark, we consider the well-performing $\text{expert}_{\text{DoW,nl}}$ model proposed in [15], extended to include two exogenous variables. We denote it by **ARX** to reflect its autoregressive structure with exogenous variables. Within this model, the transformed price on day d and hour h is given by:

$$\begin{aligned} X_{d,h} = & \underbrace{\beta_{h,1} X_{d-1,h} + \beta_{h,2} X_{d-2,h} + \beta_{h,3} X_{d-7,h}}_{\text{autoregressive effects}} + \underbrace{\beta_{h,4} X_{d-1,min} + \beta_{h,5} X_{d-1,max}}_{\text{non-linear effects}} + \underbrace{\beta_{h,6} X_{d-1,24}}_{\text{midnight value}} \\ & + \underbrace{\beta_{h,7} C_{1,d,h} + \beta_{h,8} C_{2,d,h}}_{\text{exogenous variables}} + \underbrace{\sum_{i=1}^{7} \beta_{h,8+i} D_{i,d}}_{\text{weekday dummies}} + \epsilon_{d,h}, \end{aligned} \tag{7}$$

where $X_{d-1,min}$ and $X_{d-1,max}$ are the previous day's minimum and maximum observations, $X_{d-1,24}$ is the previous day's midnight value, i.e., the last known observation in the day-ahead market, $C_{1,d,h}$ and $C_{2,d,h}$ are two exogenous variables (here: day-ahead forecasts of price drivers relevant for a given

dataset, see Section 2), $D_{1,d}, ..., D_{7,d}$ are weekday dummies, and $\epsilon_{d,h}$ is the noise term. The parameters of the model are estimated using OLS, independently for each hour.

3.4. The LEAR Model

The structure of the main predictive model considered in this study is a natural extension of Eq. (7) that includes all 24 hourly observations for the considered three days, i.e., yesterday, two days ago, and a week ago, and predictions of the exogenous variables for all 24 hours of the day:

$$\begin{aligned} X_{d,h} = &\sum_{i=1}^{24} \left(\beta_{h,i} X_{d-1,i} + \beta_{h,24+i} X_{d-2,i} + \beta_{h,48+i} X_{d-7,i}\right) + \beta_{h,73} X_{d-1,min} + \beta_{h,74} X_{d-1,max} \\ &+ \sum_{i=1}^{24} \left(\beta_{h,74+i} C_{1,d,i} + \beta_{h,98+i} C_{2,d,i}\right) + \sum_{i=1}^{7} \beta_{h,122+i} D_{i,d} + \epsilon_{d,h}. \end{aligned} \tag{8}$$

We estimate the 129 regressors via the *least absolute shrinkage and selection operator* (LASSO) [41]:

$$\widehat{\beta}_h = \arg\min_{\beta} \left(\text{RSS} + \alpha \sum_{i=1}^{129} \beta_{h,i} \right), \tag{9}$$

where RSS is the residual sum of squares and $\alpha \geq 0$ is a tuning parameter. For $\alpha = 0$ we get the standard OLS estimator, for large values of α all coefficients become zeros, while for intermediate α's the LASSO shrinks some of them to zero and thus performs variable selection. There are several techniques to optimize the tuning parameter [42]. Here, we use cross-validation with 7 folds since the number of observations in the calibration window is divisible by 7. Following [21], we refer to this *LASSO-estimated autoregressive* model as the **LEAR** model although the LEAR model defined in [21] is a richer structure (it includes nearly 250 regressors, most notably lagged fundamental variables).

3.5. Forecast Averaging Schemes

The best-performing seasonal decomposition is not known in advance. To address this problem we utilize forecast averaging [28–30]. As suggested in [8], we generate a pool of forecasts from a single model (here: ARX or LEAR) calibrated to data decomposed with different filters (either wavelet- or HP-based) and combine them into a single value. We consider two methods of combining point forecasts – one that selects the best combination of a pool of models and one inspired by Bayesian Model Averaging [27].

Moreover, since our results (see Section 4) do not provide a clear indication for the optimal order of applying transformations – seasonal decomposition first and variance stabilization second, i.e., VST[SD(·)], or variance stabilization first and seasonal decomposition second, i.e., and SD[VST(·)] – we also consider combining forecasts from both approaches. Overall, the combined predictions are constructed from individual forecasts: either 18 wavelet-based (9 smoothing levels and two orders of applying transformations) or 18 HP filter-based (9 values of λ and two orders of applying transformations).

3.5.1. Best Combination

The simpler averaging scheme selects the best performing combination of a pool of models. Hence, we refer to it as the Best Combination and denote by **BC**. Overall, given 18 individual forecasts ($\rightarrow$ columns in the left rectangle in Fig. 4), there are $2^{18} - 1$ possible combinations ($\rightarrow$ rows in the same rectangle). The first combination is just the forecast for the first LTSC, the second is the forecast for the second LTSC, ..., the 19th is the combination of forecasts for the first and second LTSCs, the 20th is the combination of forecasts for the first and third LTSCs, ..., and the last is the combination of forecasts for all LTSCs.

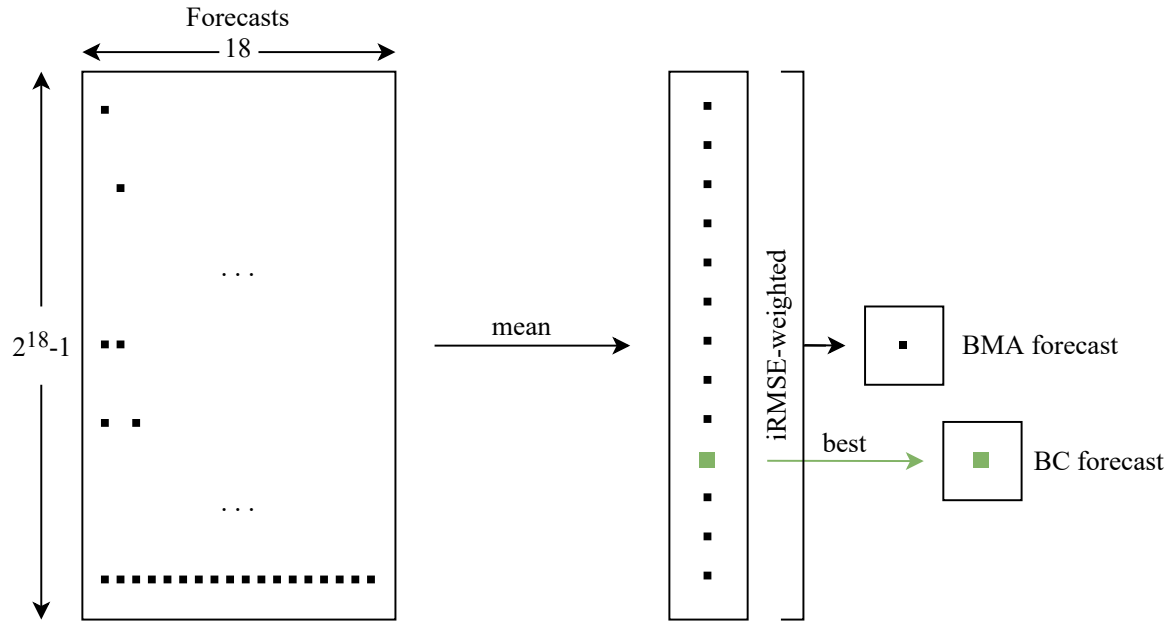

Figure 4. Illustration of the BC and BMA-type averaging schemes. For both approaches all possible combinations ($2^{18} - 1$ rows in the left rectangle) of individual forecasts for the different variants of the LTSC ($2 \times 9 = 18$ columns representing two orders of applying transformations, VST[SD(·)] and SD[VST(·)], and 9 wavelet levels or 9 HP filter's λs; see also Table 1) are generated. Then, the predictions of all models for a given combination are averaged using the arithmetic mean and either the best-performing combination in terms of RMSE in the model selection window is selected ($\rightarrow$ BC) or all combined forecasts are iRMSE-weighted ($\rightarrow$ BMA).

In each row the predictions are averaged using the arithmetic mean. Then, in order to choose the best-performing combination, we calculate the *root mean squared error* (RMSE) for all considered combinations of forecasts in the model selection window (i.e., the 2-year period directly preceding the 3-year test period, see Figs. 1 and 2):

$$\text{RMSE} = \sqrt{\frac{1}{728 \cdot 24} \sum_{d=365}^{1092} \sum_{h=1}^{24} (p_{d,h} - \hat{p}^{c}_{d,h})^2}, \tag{10}$$

where $\hat{p}^{c}_{d,h}$ is one of the $2^{18} - 1$ combined forecasts and $p_{d,h}$ is the actual price for day d and hour h. We choose the combination that returns the lowest RMSE, see the green square in Fig. 4. This combination is eventually evaluated in the model evaluation window (i.e., the 3-year test period) and compared to other combined and individual models.

In Table 1 we present the five best-preforming combinations in the model selection window obtained for both datasets and the LEAR model with wavelet-based filtering. Squares indicate individual predictions that are averaged in a given combination. For instance, for the NP dataset the best combination is composed of forecasts for three LTSCs obtained by first applying seasonal decomposition, then the VST: S_8, S_{10} and S_{13}, and one LTSC obtained by first applying the VST, then seasonal decomposition: S_{10}. The RMSE of this combination in the model selection window is 2.2984, nearly the same as that of the second best combination. The last row in each part of the table represents averaging over all individual forecasts; for NP the RMSE of 2.3485 is ca. 2% higher than for the best-performing combination.

Table 1. Root mean squared errors (RMSE) and their inverse values (iRMSE) calculated in the 2-year model selection window (see Figs. 1 and 2) for five best-performing combined forecasts obtained from the LEAR models with wavelet-based filtering. Squares indicate individual forecasts that are averaged in a given combination. For comparison, the last row represents averaging over all individual forecasts. Individual models where seasonal decomposition is executed first, then the VST is applied, are labeled by VST[SD(·)], whereas models with the opposite order are labeled by SD[VST(·)].

NP																				
	VST[SD(·)]									SD[VST(·)]										
	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	RMSE	iRMSE
1			■		■			■						■					2.2984	0.4351
2				■	■									■					2.2984	0.4351
3			■		■									■					2.2985	0.4351
4				■	■			■						■					2.2991	0.4350
5			■		■		■							■					2.3002	0.4347
	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2.3485	0.4258

PJM																				
	VST[SD(·)]									SD[VST(·)]										
	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	RMSE	iRMSE
1								■	■						■				19.0596	0.0525
2								■	■									■	19.1281	0.0523
3								■	■					■					19.1423	0.0522
4								■	■							■			19.1633	0.0522
5						■		■	■					■	■				19.2004	0.0521
	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	20.4995	0.0488

3.5.2. BMA-type Averaging

The more complex averaging scheme, instead of selecting one forecast combination out of $2^{18}-1$ possibilities, weighs all combinations:

$$\hat{p}^{\mathrm{BMA}}_{d,h} = \sum_{i=1}^{2^{18}-1} w_i \hat{p}^{c,i}_{d,h}, \tag{11}$$

where $\hat{p}^{c,i}_{d,h}$ is the ith of the $2^{18}-1$ combined forecasts. The weights are computed based on past performance:

$$w_i = \frac{\mathrm{iRMSE}_i}{\sum_{i=1}^{2^{18}-1} \mathrm{iRMSE}_i}, \tag{12}$$

where iRMSE_i is the inverse of the root mean squared error of the ith combination in Eq. (10), see Fig. 4. In the last column of Table 1 we report iRMSE values for the five best-preforming combinations obtained for both datasets and the LEAR model with wavelet-based filtering. As can be seen, due to small differences in RMSE values, the weights are nearly identical for the top performing combinations and very similar to the weight for the average across all individual forecasts.

This idea is similar in spirit to Bayesian Model Averaging [27], which weighs combinations by posterior probabilities. The latter has not performed very well in an extensive EPF study [30], hence our idea to replace posterior probabilities by iRMSE weights, similarly as in [28,29]. Although our approach is only inspired by Bayesian Model Averaging, for simplicity of notation we denote it by **BMA**.

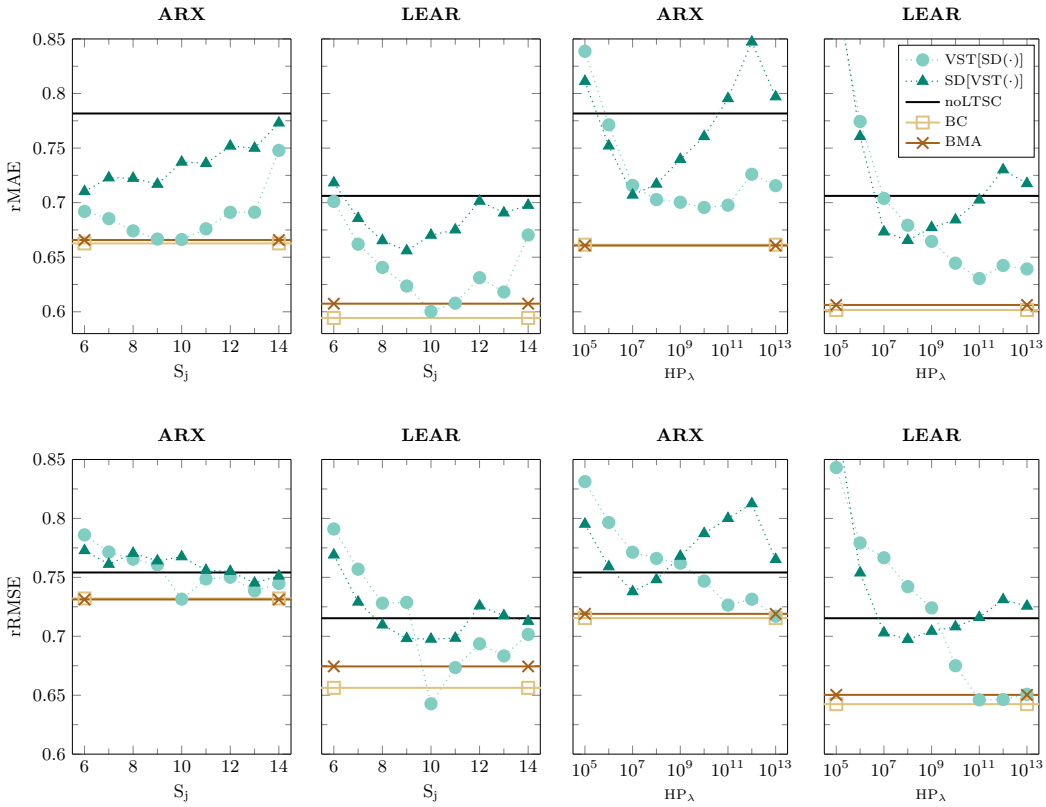

Figure 5. Relative mean absolute errors (rMAE; *top row*) and relative root mean squared errors (rRMSE; *bottom row*) for individual and combined forecasts obtained for the ARX (*left panels*) and LEAR (*right panels*) models and the NP dataset. Circles represent the performance of individual models where seasonal decomposition is executed first, then the VST is applied, while triangles the opposite order. Models without seasonal decomposition are indicated by the black solid line (labeled noLTSC), whereas the combined models by colored solid lines with markers at the ends.

4. Results

4.1. Forecast Evaluation

Forecast accuracy is assessed in the 3-year or $3 \cdot 364 = 1092$-day model evaluation window, using two error measures. Following the recommendations put forward in [21], both are relative metrics. This allows us to compare model performance across different datasets. The *relative mean absolute error* (rMAE) and the *relative root mean squared error* (rRMSE) are defined as:

$$\text{rMAE} = \frac{\sum_{d=1093}^{2184}\sum_{h=1}^{24}\left|p_{d,h}-\hat{p}_{d,h}\right|}{\sum_{d=1093}^{2184}\sum_{h=1}^{24}\left|p_{d,h}-\hat{p}^{\text{naive}}_{d,h}\right|}, \qquad \text{rRMSE} = \frac{\sqrt{\sum_{d=1093}^{2184}\sum_{h=1}^{24}\left(p_{d,h}-\hat{p}_{d,h}\right)^2}}{\sqrt{\sum_{d=1093}^{2184}\sum_{h=1}^{24}\left(p_{d,h}-\hat{p}^{\text{naive}}_{d,h}\right)^2}}, \tag{13}$$

where $p_{d,h}$ denotes the actual (observed) price for day d and hour h, $\hat{p}_{d,h}$ is the corresponding prediction obtained using the model under the evaluation, and $\hat{p}^{\text{naive}}_{d,h}$ is a similar-day prediction [43]:

$$\hat{p}^{\text{naive}}_{d,h} = \begin{cases} p_{d-7,h} & \text{for Monday, Saturday, and Sunday,} \\ p_{d-1,h} & \text{otherwise.} \end{cases} \tag{14}$$

Results for the individual and combined ARX and LEAR models for both datasets are presented in Figs. 5 and 6. They all lead to similar conclusions.

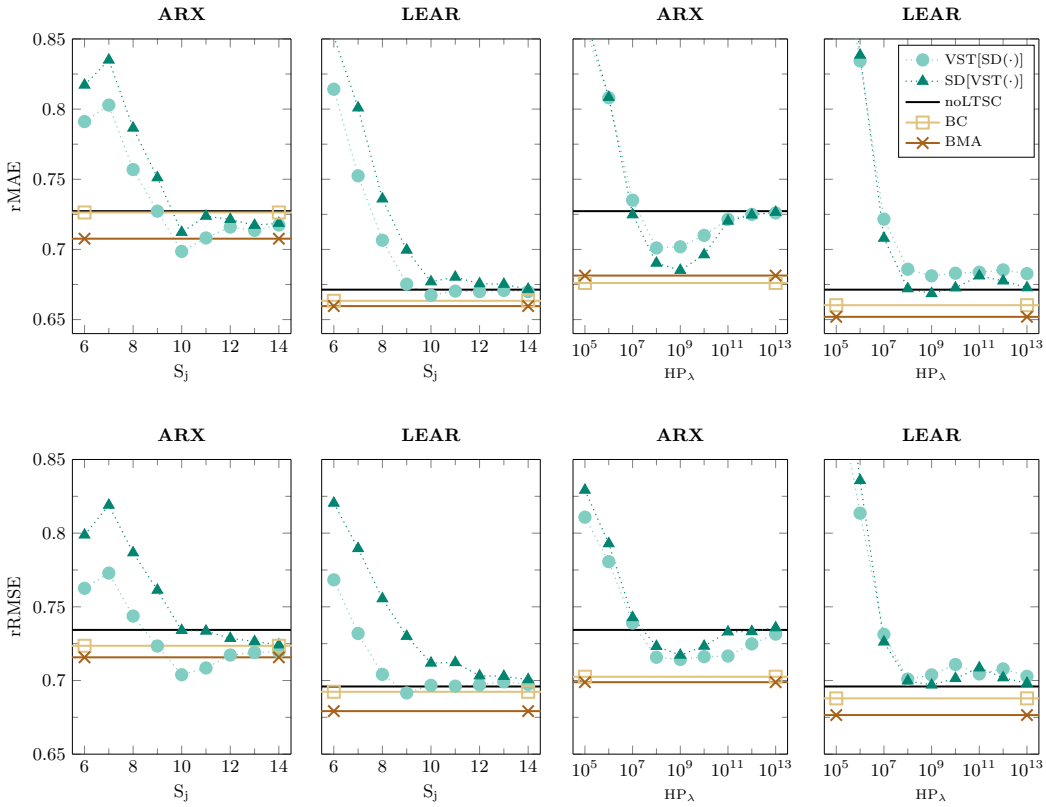

Figure 6. Relative mean absolute errors (rMAE; *top row*) and relative root mean squared errors (rRMSE; *bottom row*) for individual and combined forecasts obtained for the **ARX** (*left panels*) and **LEAR** (*right panels*) models and the PJM dataset. Markers and line styles used are the same as in Fig. 5.

Firstly, we can see that seasonal decomposition can improve the accuracy of forecasts generated not only by simple autoregressive models but also by parameter-rich models with automated variable selection via the LASSO. More considerable improvements can be achieved in the NP market, see Fig. 5. Moreover, in this case, most of the individual models with seasonal decomposition return lower error scores than the benchmark model without it (denoted by noLTSC). For the PJM dataset, see Fig. 6, the improvements are smaller, and they occur more frequently for ARX models. The only problem with such individual models is that not all of them beat the benchmark model without seasonal decomposition. Thus, it is difficult to choose the optimal LTSC ex-ante.

Secondly, there is no general answer to the question which order of applying data transformations performs better. It seems to depend on the considered model, market, type of the LTSC, or error measure. For instance, in the top row of Fig. 6, where the rMAE is presented for the PJM dataset, the best performing approach changes depending on the considered filtering type. For wavelet-based filters, it is generally more effective to apply the seasonal decomposition first, whereas for HP-based filters, the VST first. On the contrary, for the NP dataset and rRMSE (see the bottom row of Fig. 5), the performance strongly depends on the chosen parameters used to extract the LTSC, i.e., the smoothing level k or the smoothing parameter λ.

To overcome these difficulties, we consider two averaging schemes, see Sec. 3.5. In Table 2, we collect all the error scores for the combined and the benchmark models. It turns out that for both datasets the BC and BMA averaging schemes always return lower rMAE and rRMSE scores than the benchmark model without seasonal decomposition (denoted by noLTSC). Moreover, there are generally only a few individual models that can outperform the combined models. Hence, we can conclude that forecast averaging solves the problem with the ex-ante selection of the best performing LTSC and order of data transformations. Furthermore, all the combined LEAR models outperform

Table 2. Relative mean absolute errors (rMAE) and relative root mean squared errors (rRMSE) calculated in the model evaluation window for the combined and the baseline ARX and LEAR models. The lowest score in a row is emphasized in bold, independently for the ARX and LEAR models, whereas the lowest score in a row across all models has a green background.

NP								
		ARX				LEAR		
	noLTSC		BC	BMA	noLTSC		BC	BMA
rMAE	0.7817	HP	0.6615	**0.6607**	0.7062	HP	**0.6016**	0.6063
		db4	**0.6626**	0.6658		db4	**0.5944**	0.6074
rRMSE	0.7541	HP	**0.7154**	0.7190	0.7153	HP	**0.6424**	0.6503
		db4	0.7321	**0.7312**		db4	**0.6563**	0.6744
PJM								
		ARX				LEAR		
	noLTSC		BC	BMA	noLTSC		BC	BMA
rMAE	0.7273	HP	**0.6761**	0.6814	0.6713	HP	0.6603	**0.6520**
		db4	0.7263	**0.7076**		db4	0.6634	**0.6596**
rRMSE	0.7343	HP	0.7026	**0.6988**	0.6959	HP	0.6879	**0.6765**
		db4	0.7236	**0.7157**		db4	0.6923	**0.6792**

their ARX counterparts, so the use of LASSO-estimated models additionally increases the accuracy of the combined forecasts.

Finally, let us compare the averaging schemes and the ways of modeling the LTSC. It turns out that the BMA approach returns lower errors than BC in 9 out of 16 cases. When it comes to the best-performing approach for extracting the LTSC, we can observe that the one based on HP filters yields lower error scores in all the cases except one (i.e., LEAR, BC averaging, NP market; see Table 2).

4.2. Testing for Conditional Predictive Ability

The obtained forecasts are also compared using the *conditional predictive ability* (CPA) test of Giacomini and White [32]. Here, following [15], we only focus on the results of the multivariate version of the test, which for each pair of models returns a single p-value for all 24 hours. The tests were conducted separately for the absolute and squared losses.

Following [10,15,21], we present the CPA test results as chessboards with a color-coded p-value, where the color ranges from dark red (higher p-values) to dark green (lower p-values). A colored square in the chessboards indicates that the predictions of the model on the x-axis are significantly better than the predictions of the model on the y-axis. The greener this field is (the lower the p-value is) the more statistically significant the difference is. A black square means that there is no statistically significant difference in predictive accuracy at the 10% level.

In Figures 7 and 8 we illustrate the results for the absolute and squared losses, respectively. As can be seen, the combined forecasts significantly outperform the predictions of the benchmark models without seasonal decomposition in 22 out of 32 cases (at the 10% level). When we consider only the LEAR models, the improvement is statistically significant in 11 out of 16 cases. Only for the PJM market (both averaging schemes for the squared losses, see Fig. 8, and LEAR BMA db4 for linear losses, see Fig. 7) the improvements over the benchmark are not statistically significant. Moreover, the outperformance of the forecasts of the combined LEAR models over their ARX counterparts is significant in 14 out of 16 cases. Thus, the CPA tests confirm that seasonal decomposition improves the accuracy of combined forecasts generated by the LASSO-estimated models.

Regarding the method of combining forecasts. Although the BMA approach returns lower errors than the BC approach in 9 out of 16 cases (see Table 2), the CPA tests indicate that outperformance is significant only in 2 cases. On the other hand, the BC approach outperforms significantly the BMA

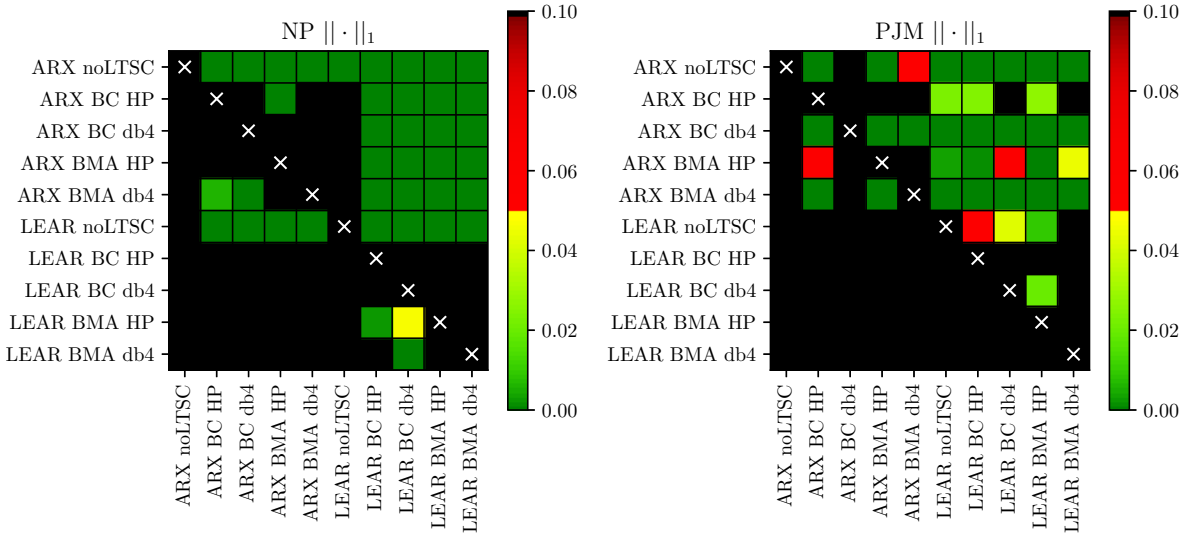

Figure 7. Results of the CPA test [32] for absolute losses. A colored square (other than black) indicates that the predictions of the model on the x-axis are significantly better than the predictions of the model on the y-axis, at a given significance level.

approach in 6 cases. Given that the differences in the error scores between both averaging schemes are small, we can recommend the simpler approach – BC averaging.

Finally, although the combined models based on HP filters return lower error scores in almost all cases, the advantages of using one LTSC function over the other are not so clear from the perspective of the CPA test. The combined forecasts based on HP filters are significantly better than those based on wavelet filters only in 6 cases, at the 10% level, whereas the remaining 10 cases show no significant differences. Considering the 5% level and only the combined LEAR models, all the differences become statistically insignificant.

4.3. Computational Complexity

The models considered in this study are computationally feasible even in time-constrained scenarios. Note, that all the times reported in this subsection reflect a single-threaded task run on an AMD Threadripper 1950X processor.

The average computation time per forecast day amounts to 8-9 s for the LEAR models and 0.03-0.05 s for the ARX models. The listed times reflect the whole process, including data preprocessing, such as the application of the LTSC or the VST. Nevertheless, model estimation is the most time-consuming task. To be more precise, the ARX model without the LTSC is, on average, computed in 0.027 s, whereas the computation of the same model with the LTSC takes 0.03 s for db4 wavelets and 0.05 s for HP filters. For the LEAR model, the input data impact the convergence rate. The smallest time in this case was measured for the model with the LTSC based on db4 wavelets – 8 s per day. On the other hand, for the model with the LTSC based on HP filters or without the LTSC, the computations took closer to 9 s.

The averaging schemes require additional computational time. One has to consider the generation of forecasts to determine the best averaging scheme. Sequential computation of the set of forecasts for one LTSC variant and a 728-day selection window takes around 33 hours. Computing the errors of all $2^{18} - 1$ possible combinations of the forecasts takes 101 seconds. The above operations are, however, done only once.

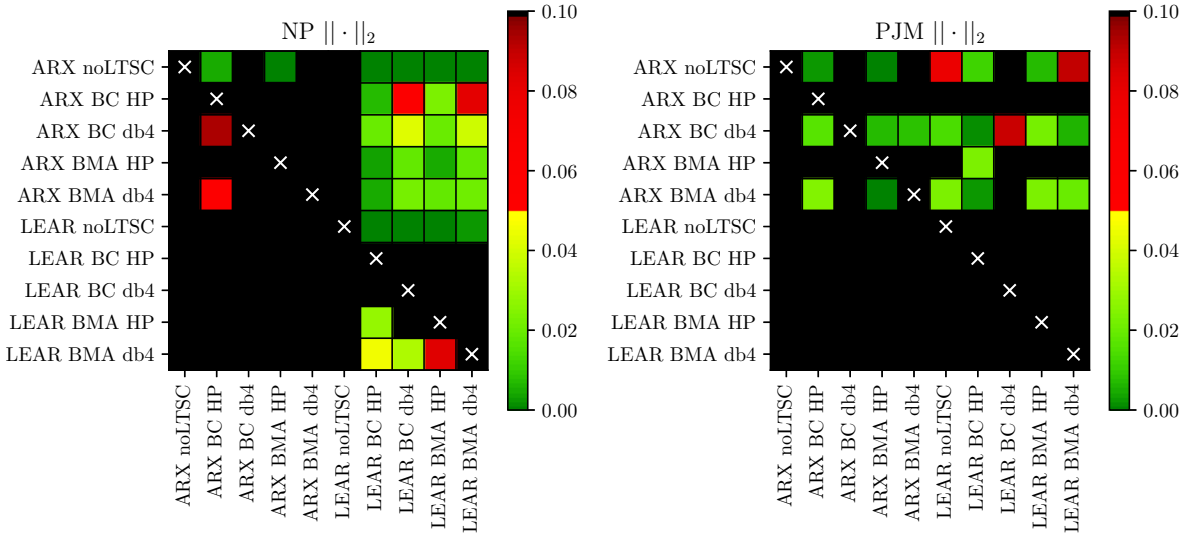

Figure 8. Results of the CPA test [32] for squared losses. A colored square (other than black) indicates that the predictions of the model on the x-axis are significantly better than the predictions of the model on the y-axis, at a given significance level.

5. Conclusions

To address the question whether the seasonal component approach is beneficial in the case of parameter-rich models, we have performed an extensive empirical study which involved a well-performing, LASSO-estimated autoregressive (LEAR) model with nearly 130 regressors, two approaches to modeling the LTSC (wavelet smoothing and the Hodrick-Prescott filter), and the area hyperbolic sine transformation. Given that our initial results did not provide a clear indication for the optimal choice of the LTSC nor the optimal order of applying transformations – seasonal decomposition first and variance stabilization second or vice versa – we have introduced two averaging schemes. The first one, dubbed Best Combination (BC), selects the best combination of a pool of models. The second is inspired by Bayesian Model Averaging, but weighs combinations by the inverse (of the) root mean squared error (iRMSE), instead of the originally proposed posterior probabilities; for notational convenience, we refer to it as BMA.

Our results indicate that seasonal decomposition can significantly – as measured by the conditional predictive ability (CPA) test – improve the accuracy of forecasts generated not only by simple autoregressive but also by parameter-rich models with automated variable selection via the LASSO. Moreover, for both datasets the averaging schemes always return lower error scores than the benchmark model without seasonal decomposition. At the same time, there are only a few individual models that can outperform the combined models. Hence, we can conclude that forecast averaging solves the problem with the ex-ante selection of the best performing LTSC and order of data transformations. Furthermore, all the combined LEAR models outperform their parsimonious ARX counterparts, so the use of LASSO-estimated models additionally increases the accuracy of the combined forecasts.

Finally, although the BMA approach generally returns slightly lower errors than BC, the CPA tests indicate that outperformance is in most cases insignificant. Hence, we recommend the simpler approach – BC averaging. Interestingly, this is consistent with recommendations put forward in the forecasting literature, i.e., that instead of combining the full set of forecasts, it may be advantageous to discard the models with the worst performance [29,44].

Author Contributions: Conceptualization, A.J, G.M, and R.W; Investigation, A.J. and G.M.; Software, A.J. and G.M.; Validation, R.W.; Writing—original draft, A.J. and G.M.; Writing—review & editing, A.J, G.M, and R.W.

Acknowledgments: This work was partially supported by the National Science Center (NCN, Poland) through MAESTRO grant No. 2018/30/A/HS4/00444 (to A.J. and R.W.) and the Ministry of Science and Higher Education (MNiSW, Poland) through Diamond Grant No. 0219/DIA/2019/48 (to G.M.).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Weron, R. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*; John Wiley & Sons, Chichester, 2006.
2. Janczura, J.; Trück, S.; Weron, R.; Wolff, R. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics* **2013**, *38*, 96–110.
3. Nowotarski, J.; Tomczyk, J.; Weron, R. Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. *Energy Economics* **2013**, *39*, 13–27.
4. Lisi, F.; Nan, F. Component estimation for electricity prices: Procedures and comparisons. *Energy Economics* **2014**, *44*, 143–159.
5. Afanasyev, D.; Fedorova, E. The long-term trends on the electricity markets: Comparison of empirical mode and wavelet decompositions. *Energy Economics* **2016**, *56*, 432–442.
6. Lisi, F.; Pelagatti, M.M. Component estimation for electricity market data: Deterministic or stochastic? *Energy Economics* **2018**, *74*, 13–37.
7. Grossi, L.; Nan, F. Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. *Technological Forecasting and Social Change* **2019**, *141*, 305–318.
8. Nowotarski, J.; Weron, R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting. *Energy Economics* **2016**, *57*, 228–235.
9. Afanasyev, D.O.; Fedorova, E.A. On the impact of outlier filtering on the electricity price forecasting accuracy. *Applied Energy* **2019**, *236*, 196 – 210.
10. Marcjasz, G.; Uniejewski, B.; Weron, R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. *International Journal of Forecasting* **2019**, *35*, 1520–1532.
11. Uniejewski, B.; Marcjasz, G.; Weron, R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting: Part II – Probabilistic forecasting. *Energy Economics* **2019**, *79*, 171–182.
12. Gaillard, P.; Goude, Y.; Nedellec, R. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting* **2016**, *32*, 1038–1050.
13. Uniejewski, B.; Nowotarski, J.; Weron, R. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* **2016**, *9*, 621.
14. Ziel, F. Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems* **2016**, *31*, 4977–4987.
15. Ziel, F.; Weron, R. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics* **2018**, *70*, 396–420.
16. Uniejewski, B.; Weron, R. Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics* **2021**, *95*, 105121.
17. Ziel, F.; Steinert, R.; Husmann, S. Efficient modeling and forecasting of electricity spot prices. *Energy Economics* **2015**, *47*, 89–111.
18. Uniejewski, B.; Weron, R. Efficient forecasting of electricity spot prices with expert and LASSO models. *Energies* **2018**, *11*, 2039.
19. Weron, R.; Zator, M. A note on using the Hodrick-Prescott filter in electricity markets. *Energy Economics* **2015**, *48*, 1–6.
20. Caldana, R.; Fusai, G.; Roncoroni, A. Electricity forward curves with thin granularity: Theory and empirical evidence in the hourly EPEXspot market. *European Journal of Operational Research* **2017**, *261*, 715–734.
21. Lago, J.; Marcjasz, G.; De Schutter, B.; Weron, R. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy* **2021**. Submitted. Working paper version available from arXiv: https://arxiv.org/abs/2008.08004.
22. Schneider, S. Power spot price models with negative prices. *Journal of Energy Markets* **2011**, *4*, 77–102.

23. Uniejewski, B.; Weron, R.; Ziel, F. Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems* **2018**, *33*, 2219–2229.
24. de Lagarde, C.M.; Lantz, F. How renewable production depresses electricity prices: Evidence from the German market. *Energy Policy* **2018**, *117*, 263–277.
25. Zhou, Y.; Scheller-Wolf, A.; Secomandi, N.; Smith, S. Managing Wind-Based Electricity Generation in the Presence of Storage and Transmission Capacity. *Production and Operations Management* **2019**, *28*, 970–989.
26. Marcjasz, G. Forecasting electricity prices using deep neural networks: A robust hyper-parameter selection scheme. *Energies* **2020**, *13*, 13184605.
27. Hoeting, J.; Madigan, D.; Raftery, A.; Volinsky, C. Bayesian model averaging: A tutorial. *Statistical Science* **1999**, *14*, 382–401.
28. Bates, J.M.; Granger, C.W.J. The Combination of Forecasts. *Operational Research Quarterly* **1969**, *20*, 451–468.
29. Bordignon, S.; Bunn, D.W.; Lisi, F.; Nan, F. Combining day-ahead forecasts for British electricity prices. *Energy Economics* **2013**, *35*, 88–103.
30. Nowotarski, J.; Raviv, E.; Trück, S.; Weron, R. An empirical comparison of alternate schemes for combining electricity spot price forecasts. *Energy Economics* **2014**, *46*, 395–412.
31. Hyndman, R.; Koehler, A. Another look at measures of forecast accuracy. *International Journal of Forecasting* **2006**, *22*, 679–688.
32. Giacomini, R.; White, H. Tests of conditional predictive ability. *Econometrica* **2006**, *74*, 1545–1578.
33. Narajewski, M.; Ziel, F. Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets* **2020**, *19*, 100107.
34. Marcjasz, G.; Uniejewski, B.; Weron, R. Beating the Naïve—Combining LASSO with Naïve Intraday Electricity Price Forecasts. *Energies* **2020**, *13*, 1667.
35. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press, 2000.
36. Conejo, A.J.; Plazas, M.A.; Espinola, R.; Molina, A.B. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems* **2005**, *20*, 1035–1042.
37. Amjady, N.; Keynia, F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy* **2009**, *34*, 46 – 57.
38. Tan, Z.; Zhang, J.; Wang, J.; Xu, J. Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy* **2010**, *87*, 3606–3610.
39. Li, S.; Wang, P.; Goel, L. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. *IEEE Transactions on Power Systems* **2015**, *31*, 1788–1798.
40. Hodrick, R.J.; Prescott, E.C. Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* **1997**, *29*, 1–16.
41. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B* **1996**, *58*, 267–288.
42. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; CRC Press, 2015.
43. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* **2014**, *30*, 1030–1081.
44. Stock, J.H.; Watson, M.W. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* **2004**, *23*, 405–430.