WORKing papers in Management Science

WORMS/20/01

Beating the naive: Combining LASSO with naive intraday electricity price forecasts

Grzegorz Marcjasz[1]
Bartosz Uniejewski[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

Article

Beating the Naïve: Combining LASSO with Naïve Intraday Electricity Price Forecasts

Grzegorz Marcjasz [1], Bartosz Uniejewski [1] and 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; Tel.: +48-71-320-4525

Version February 2, 2020 submitted to Energies

Abstract: A recent electricity price forecasting study claims that the German intraday, continuous-time market for hourly products is weak-form efficient, i.e., that the best predictor for the so-called ID3-Price index is the most recent transaction price. Here, we undermine this claim and show that we can beat the naïve forecast by combining it with a prediction of a parameter-rich model estimated using the least absolute shrinkage and selection operator (LASSO). We further argue, that that if augmented with timely predictions of fundamental variables for the coming hours, the LASSO-estimated model itself can significantly outperform the naïve forecast.

Keywords: Intraday electricity market; ID3-Price index; Price forecasting; Variable selection; Fundamental variables; LASSO; Averaging forecasts

1. Introduction

After performing a comprehensive empirical study on intraday electricity price forecasting and considering models with tens of thousands of regressors, Narajewski and Ziel [1] conclude that the German continuous-time market for hourly products is weak-form efficient, i.e., that the best predictor is the most recent transaction price. Their result is surprising and at the same time disappointing from a research perspective. Here, we undermine their claim and show that it is possible to build models that significantly outperform the naïve benchmark. Consequently, we invalidate the conjecture that the German intraday market for hourly products is weak-form efficient.

This paper belongs to a new strand of literature on forecasting prices in intraday electricity markets. To date, the workhorse of power trading in Europe has been the uniform price auction, and a vast majority of research and applications have concerned *day-ahead* (DA) electricity prices [2]. However, the rapid expansion and integration of renewable energy sources (most notably wind and solar), active demand side management (smart meters, smart appliances, etc.) and the introduction of the XBID pan-European trading platform have shifted the focus to intraday markets [3,4]. One of the more liquid – and hence more studied – marketplaces, is the German intraday market for quarter-hourly and hourly products [5–11]. In this continuous-time market, the majority of trading takes place in the last couple of hours before gate closure [12] and on the hourly products [1]; the latter are traded from 15:00 on day $d-1$ until 5 minutes before the delivery starts on day d, or 30 minutes before if the trade is made between the delivery zones. The leading reference price is the so-called *ID3-Price index* (or simply ID3), which is also an underlying instrument of exchange-traded derivative products (see www.eex.com). The index is computed as the volume-weighted average price of all trades on the quarter-hourly and hourly products in the three hour window directly preceding the delivery (see www.epexspot.com).

In this article, we focus on predicting the ID3-Price index a few hours-ahead and develop regression type models that outperform the naïve benchmark. To this end, we consider a large

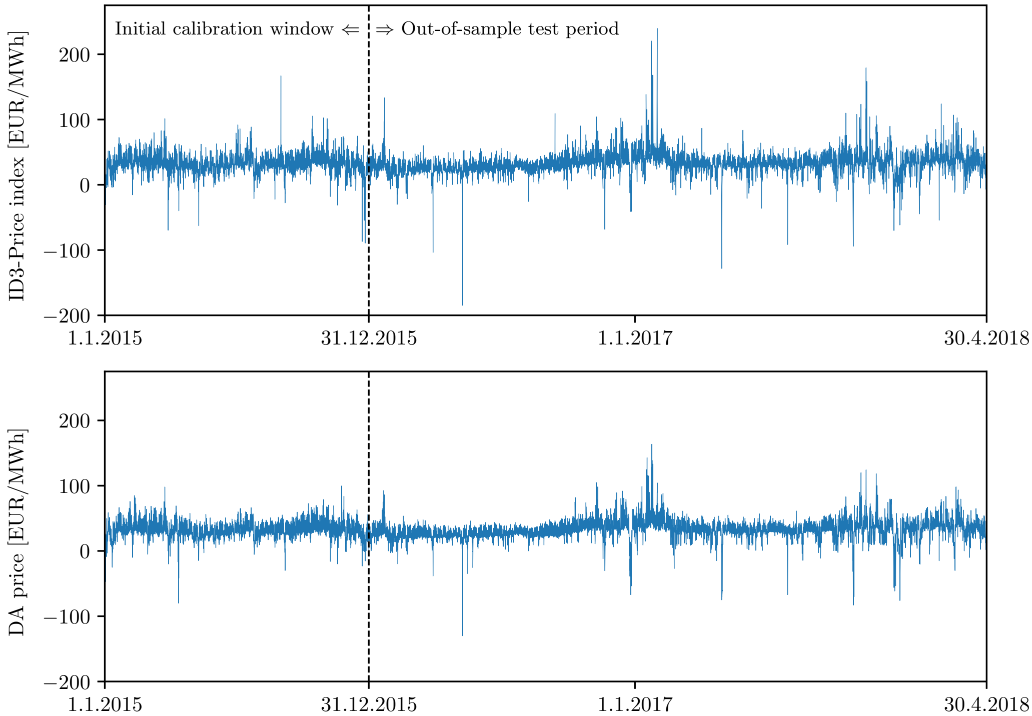

Figure 1. ID3-Price index $ID3^{d,h}$ (*top*) and day-ahead prices $DA^{d,h}$ (*bottom*) from 1.01.2015 to 30.04.2018. The vertical dashed lines mark the beginning of the 852-day long out-of-sample test period.

set of past ID3 values, past DA prices and forward-looking fundamental variables, and utilize the *least absolute shrinkage and selection operator* (LASSO) [13] to eliminate regressors with low explanatory power, as well as apply forecast averaging [14]. By comparing performance of different model structures, we draw important conclusions regarding variable selection and provide recommendations for very short-term electricity price forecasting.

The remainder of the paper is structured as follows. In Section 2, we introduce the dataset and discuss the use of variance stabilizing transformations (VSTs). Next, in Section 3 we describe the naïve approach proposed by Narajewski and Ziel [1] and introduce the model structures used in our study. In Section 4, we compare the predictive performance in terms of two commonly used error measures and the Giacomini and White [15] test for conditional predictive ability. Finally, in Section 5, we wrap up the results and conclude.

2. The Dataset

2.1. The ID3-Price Index and DA Prices

The ID3-Price index takes into account only the most recent trades, i.e., transactions that took place no earlier than 3 hours before delivery. EPEX SPOT SE publishes the index, however, the currently covered period is too short for a comprehensive evaluation of the forecasts. Therefore, following Narajewski and Ziel [1] and Uniejewski *et al.* [10], we use an ID3-like time-series reconstructed from the individual transactions and denote it by $ID3^{d,h}$, where d is the day and h is the hour of delivery, see the top panel in Figure 1. In addition to past ID3 values, we also use prices from the German day-ahead (DA) market, see the bottom panel in Figure 1. Recall, that the DA prices are set around noon on day $d-1$ for all 24 hours of day d; we denote them by $DA^{d,h}$.

Both time series are of hourly resolution and span 1216 days ranging from 1.01.2015 to 30.04.2018. Like the majority of electricity price forecasting studies, we consider a rolling window scheme. Initially,

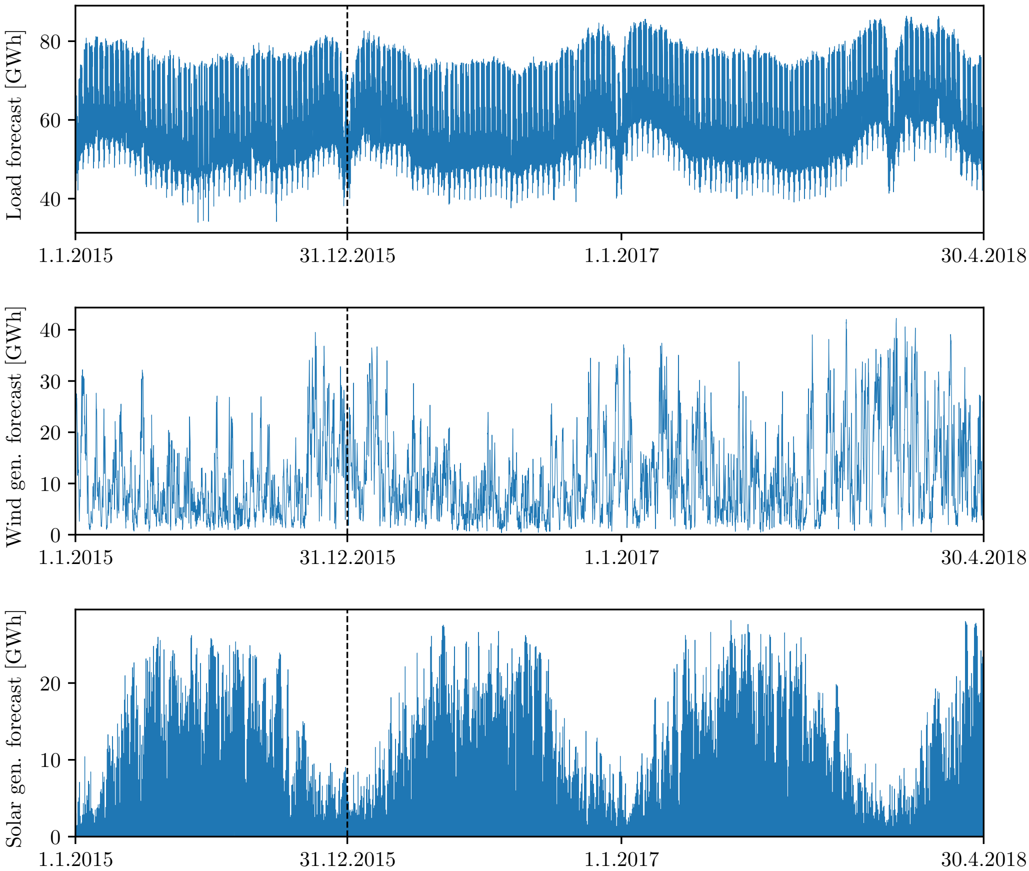

Figure 2. Three forward-looking fundamental time-series: system-wide load forecasts (*top*), wind generation forecasts (*middle*) and solar generation forecasts (*bottom*) for the period from 1.01.2015 to 30.04.2018. All three are published on day $d-1$ and concern the 24 hours of day d. As in Figure 1, the vertical dashed lines mark the beginning of the 852-day long out-of-sample test period.

we fit our models to data from 1.01.2015 hour 1 to 30.12.2015 hour 24 (i.e., we use a 364-day window) and compute the price forecasts for the first hour of 31.12.2015. Next, the window is rolled forward by 1 hour and the predictions for the second hour of 31.12.2015 are generated. This procedure is repeated until forecasts for the last hour in the 852-day long out-of-sample test period (i.e., 30.04.2018 hour 24) are made.

2.2. Exogenous Variables

The set of exogenous variables considered in this study includes three pairs of time-series that describe the demand-supply relationship in Germany:

- the system-wide load $X_1^{d,h}$ and its day-ahead forecast $\widehat{X}_1^{d,h}$,
- the total wind power generation (WPG; off- and on-shore) $X_2^{d,h}$ and its day-ahead forecast $\widehat{X}_2^{d,h}$,
- and the total photovoltaic generation (PVG) $X_3^{d,h}$ and its day-ahead forecast $\widehat{X}_3^{d,h}$,

where d is the target day and h is the hour. The day-ahead forecasts $\widehat{X}_i^{d,h}$ are plotted in Figure 2; the corresponding actual values $X_i^{d,h}$ of the fundamental variables are indistinguishable from them at this resolution. Naturally, the latter are known *ex-post*, hence only their lagged values can be used for forecasting. As discussed in Section 3, we utilize them by constructing a series of forecast errors, i.e., $\widehat{X}_i^{d,h} - X_i^{d,h}$, for the time moments for which the actual values are available; we assume that $X_i^{d,h}$ is known immediately after its hourly period ends, i.e., at $(d, h+1)$. Although an assumption, advances in on-line data collection significantly reduce the latency from the data source to the data provider, to the extent that in the near future this may become reality.

As Goodarzi *et al.* [3] argue, wind and photovoltaic generation forecasting errors increase the absolute levels of system imbalance in Germany and these in turn influence electricity prices. Hence, we additionally use a set of balancing volumes $B_i^{d,h-5}$ for the three ($i = 1, 2, 3$) quarter-hourly periods directly preceding the time at which the forecast is made, i.e., the period spans the first 45 minutes of the hour preceding the moment of computing the forecast. As in Narajewski and Ziel [1], $B_i^{d,h}$ is defined as the sum of imbalances of all German Transmission System Operators for day d and hour h; this data is published every quarter-hour, 15 minutes after the end of the delivery.

2.3. Variance Stabilizing Transformation

Following the recommendations put forward by Uniejewski *et al.* [16], we use the so-called Variance Stabilizing Transformation (VST) to reduce the impact of extreme observations present in demand, generation and particularly in electricity price data. Before applying the VST, each variable is standardized by subtracting the sample median and dividing by the sample Median Absolute Deviation (MAD) or by the sample standard deviation if MAD $= 0$, corrected by the 75th percentile of the standard normal distribution $z_{0.75}$:

$$\xi = z_{0.75} \frac{\psi - \text{Median}(\boldsymbol{\psi})}{\text{MAD}(\boldsymbol{\psi})}, \tag{1}$$

where $\boldsymbol{\psi}$ is the in-sample vector of a given variable, ψ is a single element of $\boldsymbol{\psi}$ and ξ its standardized value. Then, we use a well performing VST – the area hyperbolic sine (asinh) – on ξ. However, unlike earlier studies, we apply the VST to each variable separately due to a large number of zero-valued observations in the PVG series:

$$\phi = \text{asinh}(\xi) = \log\left(\xi + \sqrt{\xi^2 + 1}\right), \tag{2}$$

where ϕ is the VST-transformed value of ψ.

The back-transformation is more tricky. Uniejewski *et al.* [16] simply set:

$$\psi = \frac{\text{MAD}(\boldsymbol{\psi})}{z_{0.75}} \sinh(\phi) + \text{Median}(\boldsymbol{\psi}). \tag{3}$$

However, Narajewski and Ziel [1] argue that the latter is not correct since in most cases $\mathbb{E}\sinh(X) \neq \sinh(\mathbb{E}X)$. As a remedy, they propose to use the following, mathematically correct back-transformation:

$$\psi = \frac{\text{MAD}(\boldsymbol{\psi})}{z_{0.75} D} \sum_{i=1}^{D} \sinh(\phi + \varepsilon_i) + \text{Median}(\boldsymbol{\psi}), \tag{4}$$

where ε_i are in-sample residuals of the model and D is the size of the calibration window. In this study we compare model performance for both back-transformations.

3. The Models

3.1. The Naïve Benchmark

Recall, that Narajewski and Ziel [1] conclude their empirical study of intraday hourly products by stating that the market is weak-form efficient, i.e., that the best predictor is the most recent transaction price. Since we want to challenge this conjecture, as our benchmark we define:

$$\textit{naïve}^{d,h} \equiv {}_4ID_{0.25}^{d,h}, \tag{5}$$

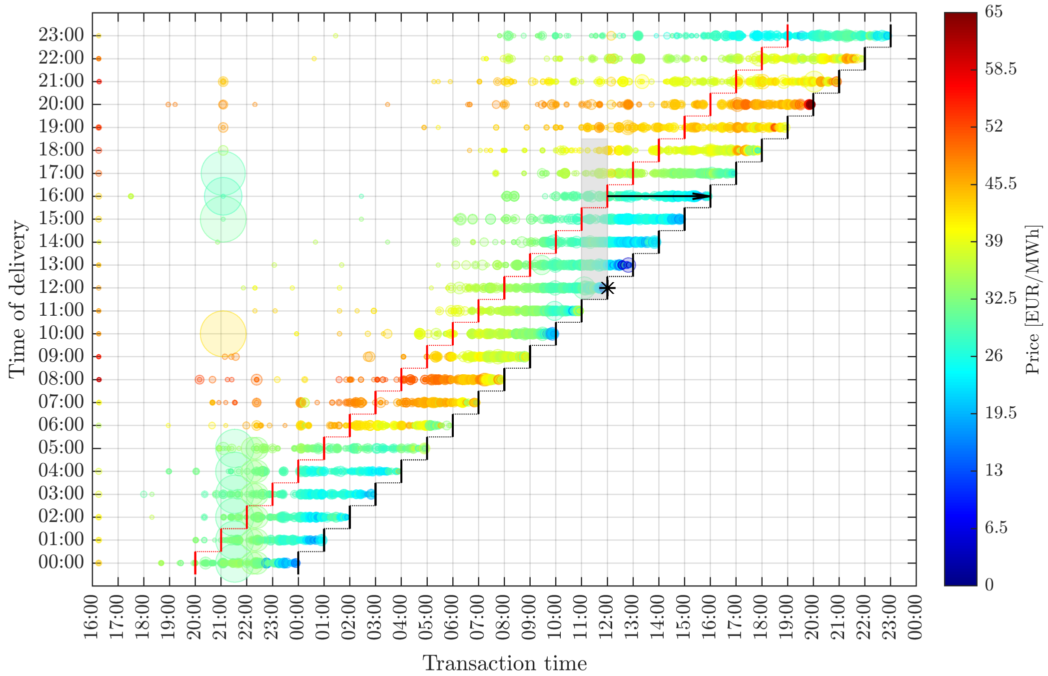

Figure 3. Illustration of the forecasting framework using actual transaction data for the period from 26.04.2018 16:00 to 27.04.2018 24:00. The black step function indicates the time the delivery starts (every hour of Friday, 27.04.2018), the circles refer to actual trades (circle size represents the traded volume – from 0.1 to 300 MWh, color represents the price – see the colorbar on the right) and the red step function represents the time the forecasts are made. For instance, at 12:00 on 27.04.2018 when forecasting the price for 16:00 ($\longrightarrow$), the most recent ID3 value is for 12:00 ($*$). The grey-shaded area indicates the data used for computing the seven partial ID3 indices utilized when forecasting the price for hour 16, see Section 3.2.3 for details.

where ${}_xID_y^{d,h}$ denotes the volume-weighted price of transactions that took place in the intraday (ID) market in a y-hour window that ended x hours before delivery on day d and hour h, see Eqn. (2) in [1]. Using this notation the ID3-Price index can be defined as $ID3^{d,h} \equiv {}_0ID_3^{d,h}$, i.e., the volume-weighted price of transactions that took place in the last three hours of trading (excluding the last 5 or 30 minutes, see Section 2).

Note, that our *naïve* benchmark is not identical to the one used in [1], i.e., Naive.MR1 $\equiv {}_{3.25}ID_{0.25}^{d,h}$. Instead of assuming that the trader makes the decision and places orders in a 15-minute window ending 3 hours before delivery, we give her a one hour window for making the trading decisions (between 4 and 3 hours before delivery). This is illustrated in Figure 3, where the red step function represents the time the forecasts are made (4 hours before delivery) and the black step function the time the delivery starts.

3.2. LASSO-estimated Models

An advantage of using automated variable selection is an almost unlimited number of initially considered explanatory variables [17]. In this study, we define a baseline model with 76 potential regressors and its three extensions; the largest one takes into account 200+ explanatory variables. All considered models are estimated in a multivariate modeling framework in the sense of Ziel and Weron [18], i.e., an explicit 'day $\times$ hour' matrix-like structure is used for the 24-dimensional price vectors. However, unlike when forecasting in day-ahead auction markets, where the prices are set once a day, in a continuous-time intraday market we are able to use information updated in the course of the day, e.g., more recent weather forecasts.

3.2.1. The Baseline Model

The baseline model is a slightly modified LASSO-estimated model of Uniejewski *et al.* [10]. The only difference is the omission of some of the less important variables. Namely, we exclude the information about inputs distant in time and only use the latest information about past ID3 and DA prices. As a result, we obtain a model with 76 potential regressors – 21 last known ID3-Price index values from the intraday market (i.e., nearly the whole day), 24 DA prices for the target day and seven dummy variables (to account for the weekly seasonality). Given the 4-hour forecast to delivery lag and the time the DA prices are published, we can additionally include next day's DA prices when forecasting hours 16 to 24:

$$ID3^{d,h} = \underbrace{\sum_{i=4}^{24} \beta_{i-3} ID3^{d,h-i}}_{\text{past intraday prices}} + \underbrace{\sum_{i=1}^{24} \beta_{21+i} DA^{d,i}}_{\text{DA prices for day } d} + \underbrace{\sum_{i=1}^{7} \beta_{45+i} D_i}_{\text{weekday dummies}} + \underbrace{\mathbb{1}_{h\geq 16} \sum_{i=1}^{24} \beta_{52+i} DA^{d+1,i}}_{\text{DA prices for day } d+1} + \varepsilon^{d,h}, \quad (6)$$

where $\varepsilon^{d,h}$ is the noise term. To simplify the notation when referring to an hourly product with delivery i hours after ($i > 0$) or before ($i < 0$) the product with delivery on day d and hour h (more precisely: with delivery between hour $h-1$ and h) we define:

$$(d, h+i) \equiv \left(d + \left\lfloor \frac{h+i-1}{24} \right\rfloor, \; h+i-24 \left\lfloor \frac{h+i-1}{24} \right\rfloor \right). \quad (7)$$

For instance, for $h = 2$ and $i = -5$ we have $(d, -3) \equiv (d-1, 21)$, while for $h = 2$ and $i = -2$ we have $(d, 0) \equiv (d-1, 24)$. Note, that the price for each hour is predicted 4 hours in advance, hence the first sum in the above formula starts with $i = 4$, and using the most recent information available, see Figure 3. Later in the text we denote model (6) by **baseline**.

3.2.2. The Model with Exogenous Variables

The first extension of model (6) is motivated by the results of Uniejewski *et al.* [19], who showed that fundamental variables play an important role when forecasting DA prices. On the other hand, Monteiro *et al.* [20] and Andrade *et al.* [21] argued that fundamentals (historical and predicted demand, generation and weather) did not have much explanatory power when forecasting Spanish intraday prices, since DA prices already included this information. To check whether fundamentals can help in forecasting the ID3-Price index in the German intraday market, we extend the baseline model to include load, wind power generation (WPG) and photovoltaic generation (PVG) forecasts and the corresponding errors, as well as the balancing volumes (Section 2.2 for details):

$$\begin{aligned} ID3^{d,h} = & \sum_{i=4}^{24} \beta_{i-3} ID3^{d,h-i} + \sum_{i=1}^{24} \beta_{21+i} DA^{d,i} + \sum_{i=1}^{7} \beta_{45+i} D_i + \mathbb{1}_{h\geq 16} \sum_{i=1}^{24} \beta_{52+i} DA^{d+1,i} + \\ & + \underbrace{\sum_{i=1}^{24} \beta_{76+i} \widehat{X}_1^{d,i}}_{\text{load forecasts}} + \underbrace{\sum_{i=1}^{24} \beta_{100+i} \widehat{X}_2^{d,i}}_{\text{WPG forecasts}} + \underbrace{\sum_{i=1}^{24} \beta_{124+i} \widehat{X}_3^{d,i}}_{\text{PVG forecasts}} + \underbrace{\sum_{i=4}^{24} \beta_{148+i} \left(\widehat{X}_1^{d,h-i} - X_1^{d,h-i} \right)}_{\text{errors of load forecasts}} + \\ & + \underbrace{\sum_{i=4}^{24} \beta_{169+i} \left(\widehat{X}_2^{d,h-i} - X_2^{d,h-i} \right)}_{\text{errors of WPG forecasts}} + \underbrace{\sum_{i=4}^{24} \beta_{190+i} \left(\widehat{X}_3^{d,h-i} - X_3^{d,h-i} \right)}_{\text{errors of PVG forecasts}} + \underbrace{\sum_{i=1}^{3} \beta_{211+i} B_i^{d,h-5}}_{\text{balancing volumes}} + \varepsilon^{d,h}. \end{aligned} \quad (8)$$

Later in the text we denote this model by **w/exogenous**.

3.2.3. The Model with Partial ID Prices

The second extension of model (6) is motivated by the results of Narajewski and Ziel [1]. The authors emphasize that the most important information for forecasting ID3 can be derived from recent transaction data for a given product. Hence, we extend the baseline model to include 8 additional predictors. Firstly, we add the *naïve* benchmark (5) as one of the explanatory variables. Secondly, we add variables that link the intraday to day-ahead markets and reflect changes in the expectations about price levels over time. More precisely, we construct artificial series that utilize the information from recent transaction data on the neighboring products. For $i = -4, ..., 2$, we define seven *partial ID indexes*:

$$pID_i^{d,h} \equiv \frac{1}{\sum_{\tau \in [(d,h-5),(d,h-4)]} V_\tau^{d,h+i}} \sum_{\tau \in [(d,h-5),(d,h-4)]} V_\tau^{d,h+i} P_\tau^{d,h+i}, \tag{9}$$

where $V_\tau^{d,h}$ and $P_\tau^{d,h}$ are respectively the volume and price of a transaction made at time τ on a product with delivery on day d and hour h. Hence, $pID_i^{d,h}$ is a volume-weighted price of all transactions on product $(d, h+i)$ in the last hour before the forecast is computed, i.e., between 5 and 4 hours before the delivery. For example, to compute $pID_i^{d,16}$, we use seven hourly windows corresponding to $i = -4, ..., 2$, see the gray-shaded rectangle spanning 7 hourly products in Figure 3. Note, that using the ${}_xID_y^{d,h}$ notation we can write:

$$pID_i^{d,h} \equiv {}_{4+i}\mathrm{ID}_1^{d,h+i}. \tag{10}$$

Finally, we can define the model with partial ID prices as follows (later in the text we denote it by **w/partial ID**):

$$\begin{aligned} ID3^{d,h} = {} & \sum_{i=4}^{24} \beta_{i-3} ID3^{d,h-i} + \sum_{i=1}^{24} \beta_{21+i} DA^{d,i} + \sum_{i=1}^{7} \beta_{45+i} D_i + \mathbb{1}_{h \geq 16} \sum_{i=1}^{24} \beta_{52+i} DA^{d+1,i} + \\ & + \underbrace{\sum_{i=-4}^{2} \beta_{81+i} \left(DA^{d,h+i} - pID_i^{d,h} \right)}_{\text{difference between DA and partial ID prices}} + \underbrace{\beta_{84}\, na\ddot{\imath}ve^{d,h}}_{\text{naïve benchmark}} + \varepsilon^{d,h}. \end{aligned} \tag{11}$$

3.2.4. The Full Model

Now, we are ready to write the full model (denoted later in the text by **full**), which includes all elements of models (8) and (11). We end up with a maximum of 222 potential regressors, depending on whether we already know the day-ahead prices for day $d+1$:

$$\begin{aligned} ID3^{d,h} = {} & \sum_{i=4}^{24} \beta_{i-3} ID3^{d,h-i} + \sum_{i=1}^{24} \beta_{21+i} DA^{d,i} + \sum_{i=1}^{7} \beta_{45+i} D_i + \mathbb{1}_{h \geq 16} \sum_{i=1}^{24} \beta_{52+i} DA^{d+1,i} + \\ & + \sum_{i=1}^{24} \beta_{76+i} \widehat{X}_1^{d,i} + \sum_{i=1}^{24} \beta_{100+i} \widehat{X}_2^{d,i} + \sum_{i=1}^{24} \beta_{124+i} \widehat{X}_3^{d,i} + \sum_{i=4}^{24} \beta_{148+i} \left(\widehat{X}_1^{d,h-i} - X_1^{d,h-i} \right) + \\ & + \sum_{i=4}^{24} \beta_{169+i} \left(\widehat{X}_2^{d,h-i} - X_2^{d,h-i} \right) + \sum_{i=4}^{24} \beta_{190+i} \left(\widehat{X}_3^{d,h-i} - X_3^{d,h-i} \right) + \sum_{i=1}^{3} \beta_{211+i} B_i^{d,h-5} + \\ & + \sum_{i=-4}^{2} \beta_{219+i} \left(DA^{d,h+i} - pID_i^{d,h} \right) + \beta_{222}\, na\ddot{\imath}ve^{d,h} + \varepsilon^{d,h} \end{aligned} \tag{12}$$

The final modification of the benchmark model is obtained by fixing $\beta_{222} \equiv 1$, as considered in [1]. Later in the text we denote such a model by **full-diff**, because it corresponds to setting the dependent variable to the difference between ID3 and the *naïve* benchmark, instead of the ID3-Price index itself.

3.3. LASSO Estimation

In order to explain the estimation scheme, let us use a more compact form of the regression model:

$$X^{d,h} = \sum_{i=1}^{n} \beta_i V_i^{d,h}, \tag{13}$$

where $V_i^{d,h}$'s are the predictors and β_i's are the corresponding coefficients. The *least absolute shrinkage and selection operator* (LASSO) shrinks the coefficients of the less important explanatory variables towards zero and hence performs variable selection [13,22]. The LASSO can be treated as a generalization of linear regression, where instead of minimizing only the *residual sum of squares* (RSS), the sum of RSS and a linear penalty function of the β's is minimized:

$$\hat{\boldsymbol{\beta}}_L = \min_{\boldsymbol{\beta}} \{\text{RSS} + \lambda \, \|\boldsymbol{\beta}\|_1\} = \min_{\boldsymbol{\beta}} \left\{ \text{RSS} + \lambda \sum_{i=1}^{n} |\beta_i| \right\}, \tag{14}$$

where $\lambda \geq 0$ is a *tuning* (or *regularization*) parameter. Note that for $\lambda = 0$, we get the standard least squares estimator, for $\lambda \to \infty$, all β_i's tend to zero, while for intermediate values of λ, there is a balance between minimizing the RSS and shrinking the coefficients.

Selecting a 'good' value for λ is critical. It is, however, a complex problem [8,11,17]. Because of a relatively short dataset, we are not able to reselect λ based on model performance in a validation period. Instead, we have decided to use cross-validation. It can be effectively applied to select the tuning parameter *ex-ante*, unfortunately at a cost of increased computational complexity. The procedure is discussed in more detail in Section 5.2.

3.4. Forecast Averaging

Combining forecasts in order to obtain more precise and robust predictions is a technique known both in the electricity price forecasting literature [14] and in forecasting in general [23]. Here, we use an arithmetic average of two predictions – obtained from the LASSO-estimated model (labeled Z) and the *naïve* forecast:

$$\text{ens}(Z) = \frac{1}{2}\, \widehat{ID3}_Z^{d,h} + \frac{1}{2}\, \textit{naïve}^{d,h}. \tag{15}$$

The motivation for using the arithmetic mean is twofold. Firstly, it is the simplest averaging scheme, requiring no additional calibration. Secondly, it is hard to beat by 'more sophisticated' approaches [24].

4. Results

4.1. Forecast Evaluation

The forecasting accuracy is assessed in terms of two error measures: the *Mean Absolute Error* (MAE) and the *Root Mean Squared Error* (RMSE). The scores are reported for the full out-of-sample test period of $D = 852$ days, i.e., 31.12.2015 to 30.04.2018, see Figure 1, jointly for all hours of the day:

$$\text{MAE} = \frac{1}{24D} \sum_{d=1}^{D} \sum_{h=1}^{24} \left| \mathcal{E}_Z^{d,h} \right| \quad \text{and} \quad \text{RMSE} = \sqrt{\frac{1}{24D} \sum_{d=1}^{D} \sum_{h=1}^{24} \left| \mathcal{E}_Z^{d,h} \right|^2}, \tag{16}$$

Table 1. MAE and RMSE errors for all 852 days of the out-of-sample test period, see Figure 1. The upper part of the table reports on the results obtained for models which use back-transformation (4), while the lower that use back-transformation (3). Columns labeled **Model** refer to the models themselves, while those labeled **ens(Model)** to ensembles with the *naïve* benchmark, as defined in Eqn. (15). Errors smaller than those of the *naïve* benchmark are emphasized in bold.

Back-transformation	Model class	MAE		RMSE	
		Model	ens(Model)	Model	ens(Model)
	naïve	3.774	—	5.999	—
With correction proposed in [1], see Eqn. (4)	baseline	4.433	3.866	7.178	6.246
	w/exogenous	4.234	**3.720**	6.956	6.040
	w/partial ID	**3.771**	**3.702**	6.052	**5.903**
	full	**3.710**	**3.630**	6.072	**5.841**
	full-diff	**3.723**	**3.700**	**5.900**	**5.906**
As originally introduced in [16], see Eqn. (3)	baseline	4.427	3.868	7.294	6.285
	w/exogenous	4.242	**3.724**	7.069	6.086
	w/partial ID	3.807	**3.708**	6.182	**5.942**
	full	**3.733**	**3.635**	6.178	**5.877**
	full-diff	**3.699**	**3.710**	**5.894**	**5.923**

where $\mathcal{E}_Z^{d,h} = ID3^{d,h} - \widehat{ID3}^{d,h}$ is the prediction error for model Z, for day d and hour h. Recall, that the RMSE is the optimal measure for least square problems, whereas the MAE is more robust to outliers [22]. The resulting aggregate MAE and RMSE scores can be used to provide a ranking of the models, but do not allow to draw statistically significant conclusions on the relative performance. Therefore, we use the Giacomini and White [15] test for *conditional predictive ability* (CPA), which can be regarded as a generalization of the commonly used Diebold-Mariano test for *unconditional* predictive ability [2]. First, for each pair of models, we compute the so-called multivariate loss differential series [16,18]:

$$\Delta_{A,B}^d = \|\widehat{\mathcal{E}}_A^d\|_p - \|\widehat{\mathcal{E}}_B^d\|_p, \tag{17}$$

where $\|\mathcal{E}_Z^d\|_p = (\sum_{h=1}^{24} |\mathcal{E}_Z^{d,h}|^p)^{1/p}$ is the p-th norm of the 24-dimensional vector of out-of-sample errors for model Z. Then, we calculate the p-values of the CPA test with null $H_0 : \boldsymbol{\alpha} = 0$ in the regression: $\Delta_{A,B}^d = \boldsymbol{\alpha}'\mathbb{X}^{d-1} + \varepsilon^d$, where $\mathbb{X}^{d-1}$ contains information for day $d-1$, i.e., a constant and lags of $\Delta_{A,B}^d$.

4.2. MAE and RMSE Errors

In Table 1 we report the MAE and RMSE metrics for all considered models and their ensembles with the *naïve* benchmark, as defined in Eqn. (15). In Figure 4 we additionally visualize the set of results corresponding to back-transformation (4), reflecting the upper part of Table 1. Several important conclusions can be drawn:

- In terms of the MAE, three models outperform the *naïve* benchmark even without averaging forecasts. However, only the **full-diff** approach manages to beat the benchmark in terms of the RMSE, see the values emphasized in bold in Table 1 in columns labeled **Model**.
- All baseline model extensions yield lower errors than the baseline model itself, both in terms of the MAE and RMSE.
- The **full** model outperforms the model with partial ID prices, which suggests that using the exogenous variables discussed in Section 2.2 improves forecast accuracy.
- On average, back-transformation (4) proposed by Narajewski and Ziel [1] (the upper part of Table 1) performs slightly better than the originally introduced one (the lower part of Table 1). For this reason, in what follows we only discuss the results of back-transformation (4).
- Apart from the **full-diff** model, every other model performs better when its forecasts are averaged using Eqn. (15). Compare the columns labeled **Model** and **ens(Model)** in Table 1.

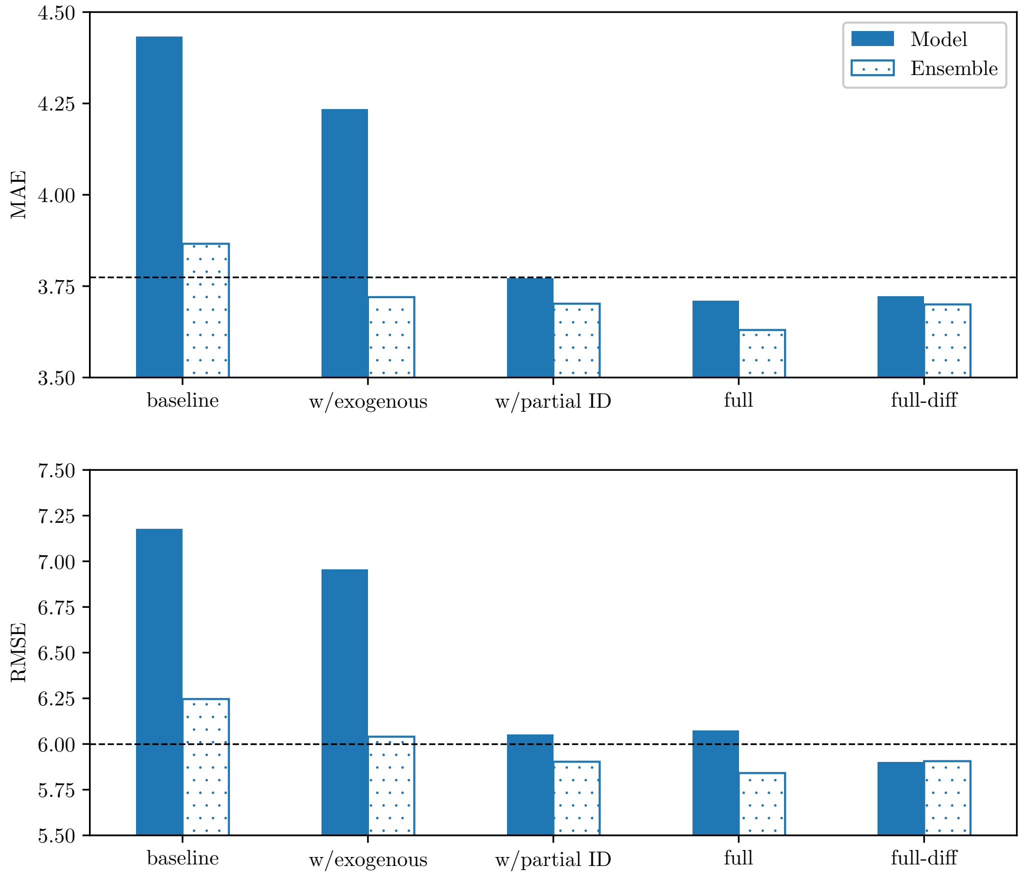

Figure 4. Bar plots illustrating the MAE (*top*) and RMSE (*bottom*) errors reported in the upper part of Table 1, i.e., for the *naïve* benchmark and models that utilize back-transformation (4). The brown dashed lines correspond to the benchmark, the solid bars represent the individual models and the dotted bars the corresponding ensembles.

- The improvements from averaging forecasts are much higher (ca. 12-14%) for models that do not use the *naïve* benchmark as a regressor. However, what is surprising, the gains are noticeable (ca. 2-4%) even for models which include this explanatory variable. Apparently, the LASSO scheme does not put enough weight to this variable. Setting $\beta_{222} = 0$ in the **full-diff** model helps, but does not solve the problem completely. We return to this issue in Section 4.4.

4.3. Conditional Predictive Ability

We perform the Giacomini and White [15] test of *conditional predictive ability* (CPA) to check whether the differences in forecasting accuracy are statistically significant. We conduct the test only for the *naïve* benchmark and models that utilize back-transformation (4). The *p*-values of the pairwise comparisons are visualized in Figure 5. We can see that:

- *Naïve* forecasts can be significantly outperformed by predictions of models that include partial ID information and exogenous variables (**full** and **full-diff** models) without averaging, and by most of models after ensembling.
- Forecasts of the baseline model are significantly outperformed by those of any other LASSO-estimated model.
- For all considered models, ensembling significantly improves the accuracy in terms of the linear errors.

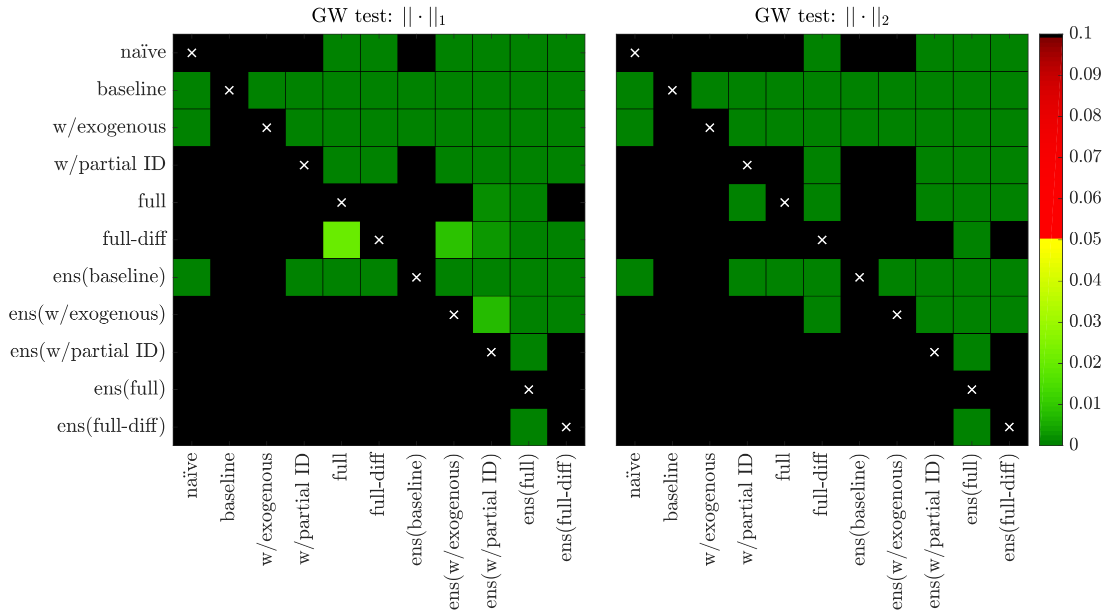

Figure 5. Results of the conditional predictive ability (CPA) test of Giacomini and White [15] for the linear (*left*) and quadratic (*right*) errors. We use a heat map to indicate the range of the *p*-values – the closer they are to zero (→ dark green) the more significant is the difference between the forecasts of a model on the X-axis (better) and the forecasts of a model on the Y-axis (worse).

- Forecasts of the **ens(full)** model significantly outperform those of any other model, both in terms of the linear and quadratic errors.

4.4. Why Does Ensembling Improve the Results?

As the above reported results indicate, the ensemble is in most cases able to outperform both individual forecasts. However, the simple averaging scheme proposed in Eqn. (15) might not be the optimal for this task. Hence, in this Section we consider a more general formula:

$$\mathrm{ens}(Z) = (1 - w) \cdot \widehat{ID3}_Z^{d,h} + w \cdot \textit{naïve}^{d,h}, \tag{18}$$

where w is the weight assigned to the *naïve* forecast. In Figure 6 we depict the MAE of ensemble (18) as a function of w for the **full** model with back-transformation (4). The MAE curve is convex with a

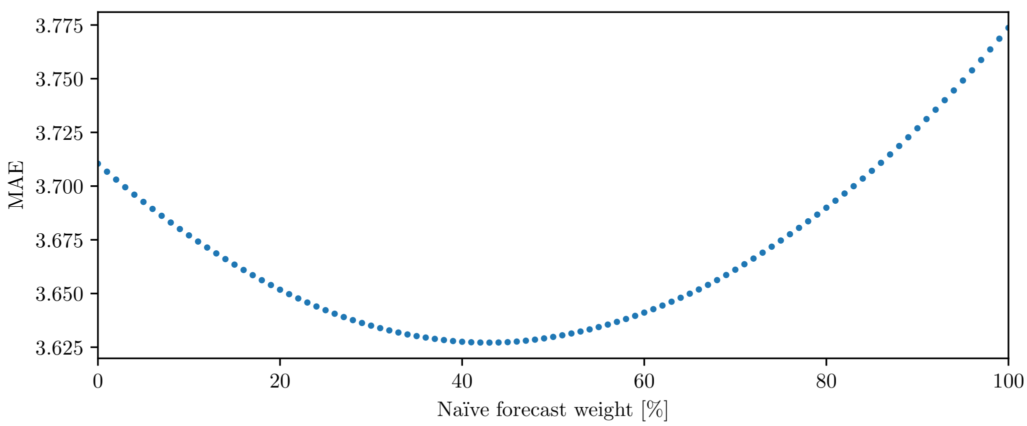

Figure 6. The MAE errors of ensembles created using Eqn. (18) that utilize the **full** model with back-transformation (4), as a function of the weight assigned to the *naïve* benchmark.

Table 2. The MAE errors in three price regimes (*from top to bottom*): the highest 2.5%, the middle 50% and the lowest 2.5% observations in the out-of-sample test period. The LASSO model used is the **full** model with back-transformation (4).

Regime	Full model	Naïve benchmark	Ensemble
High	12.418	11.938	12.025
Middle	2.675	2.795	2.631
Low	13.957	12.352	12.859

minimum at ca. $w = 45\%$. However, the value for $w = 50\%$, i.e., the simple mean used in the study, is very close to the optimum.

The reason behind this shape is the characteristic of LASSO forecasts, estimated on long calibration windows. Specifically, the model is trained to generalize well, and such a behavior is reinforced by the fact that there are only a few spikes in the calibration window. As such, the model is able to better predict prices at the typically observed levels at the cost of underestimating spikes (both positive and negative), see Table 2. Therefore the ensemble (regardless of the weights) balances the generalization of the LASSO forecasts with the ability to quickly adapt to non-recurring phenomena of the *naïve* benchmark.

5. Discussion and Conclusions

The motivation for this study was a claim made by Narajewski and Ziel [1], that the German intraday, continuous-time market for hourly products was weak-form efficient, i.e., that the best predictor for the ID3-Price index was the most recent transaction price. Performing a comprehensive forecasting exercise involving parameter-rich regression-type models with four types of fundamental variables as inputs, we have been able to challenge their claim and show that we can significantly outperform the *naïve* forecast by combining it with a prediction of a LASSO-estimated model. To keep the empirical part of the paper concise, we have opted for omitting some of the considerations. Let us now briefly discuss them.

5.1. *The Moment of Forecasting the ID3-Price Index*

After consulting with practitioners, we have decided to focus on a forecasting scheme used by Uniejewski *et al.* [10], where the predictions are made four hours before delivery. This means, that a trader has an hour to make the decisions and build a long or short position before the ID3 transaction window opens three hours before delivery. However, to check whether also the Naive.MR1 $\equiv {}_{3.25}ID^{d,h}_{0.25}$ benchmark of Narajewski and Ziel [1] can be outperformed, we have recalculated our models in their setting. Naturally, the Naive.MR1 is harder to beat than our *naïve* model, because it uses more recent transaction data. Yet, the relative performance vs. the benchmark was qualitatively the same as reported in Section 4.

5.2. *Selecting the LASSO Regularization Parameter*

For the choice of the regularization parameter, we have resorted to using an automated cross validation (CV) technique. More precisely, the applied CV procedure consisted of three folds with a dense logarithmic grid of 50 λ values spanning six orders of magnitude. Two thirds of the calibration sample was used for training the models estimated with different λ's, the remaining one third for testing them. This resulted in a significantly increased computational burden, due to the need of testing multiple models for multiple λ's, but also allowed for an *ex-ante* choice of the regularization parameter. We have also performed a limited numerical experiment to compare with the results obtained for the best *ex-post* selected λ. As it turned out, the difference in the MAE and RMSE errors was less than 0.5%.

5.3. The Impact of Intraday Updates of the Fundamentals

We have also tried to assess the impact of using more recent forecasts of the system-wide load, wind power generation, photovoltaic generation and balancing volumes. We have measured the predictive performance of our models under the assumption that we know future values of the exogenous variables until the end of the target day. With such 'perfect forecasts' we have been able to additionally reduce the forecasting error by more than 2%. This result emphasizes how important in short-term forecasting is the availability of more frequently updated forecasts of the exogenous variables.

5.4. Model Size

As mentioned above, the LASSO procedure allows for an efficient estimation of parameter-rich models. However, the quality of the obtained estimates can differ for different sizes of the regression model. Having only ca. 360 observations in the calibration window, we may obtain worse forecasts if we consider dozens or hundreds of redundant variables in the model. The **full** model defined by Eqn. (12) includes only ca. 200 potential predictors. Interestingly it outperforms by ca. 0.6% a richer model with more than 800 variables (the same information sources, but more past observations). Therefore we advise to use expert knowledge and/or back-testing to eliminate non-informative predictors before running the LASSO.

5.5. Directions for Future Research

Given that the literature on forecasting prices in European intraday power markets is still very scarce, our study is a step forward towards understanding the impact of using recent transaction data and exogenous variables on the predictive performance. Our study can be further expanded in several directions. In particular, we report the results for only one VST (for more suggestions see [16]) and without decomposing the data into a long-term seasonal component and the remaining stochastic part (for the importance of doing this see, e.g., [25,26]). Furthermore, we have focused on point forecasting, ignoring the full predictive distribution [7,27] or – what may be even more important in continuous-time intraday markets – the trajectories [12,28]. We have restricted ourselves to using regression-based models, however, machine learning techniques could be used in this context as well [11,20,21,29], naturally at the cost of an increased computational burden. Finally, recall from Section 4.4, that the ensemble we use balances the generalization of the LASSO forecasts with the ability to quickly adapt to non-recurring phenomena of the *naïve* benchmark. A potentially viable alternative would be to use the approach introduced by Hubicka *et al.* [30], which averages forecasts of a given model across calibration windows of different length.

Author Contributions: Conceptualization, R.W.; investigation, G.M. and B.U.; software, G.M. and B.U.; validation, all authors; writing—original draft, G.M. and B.U.; writing—review and editing, all authors. All authors have read and agree to the published version of the manuscript.

Funding: This work was partially supported by the National Science Center (NCN, Poland) through grant No. 2018/30/A/HS4/00444 (to G.M.), the Ministry of Science and Higher Education (MNiSW, Poland) through grant No. 0199/DIA/2019/48 (to B.U.), and the German Research Foundation (DFG, Germany) and the National Science Center (NCN, Poland) through grant No. 2016/23/G/HS4/01005 (to R.W.).

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

References

1. Narajewski, M.; Ziel, F. Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets* **2019**. (DOI: 10.1016/j.jcomm.2019.100107).
2. 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.
3. Goodarzi, S.; Perera, H.; Bunn, D. The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices. *Energy Policy* **2019**, *134*, 110827.

4. Kath, C. Modeling intraday markets under the new advances of the cross-border intraday project (XBID): Evidence from the German intraday market. *Energies* **2019**, *12*, 4339.
5. Ciarreta, A.; Muniain, P.; Zarraga, A. Modeling and forecasting realized volatility in German–Austrian continuous intraday electricity prices. *Journal of Forecasting* **2017**, *36*, 680–690.
6. Kiesel, R.; Paraschiv, F. Econometric analysis of 15-minute intraday electricity prices. *Energy Economics* **2017**, *64*, 77–90.
7. Bunn, D.; Gianfreda, A.; Kermer, S. A trading-based evaluation of density forecasts in a real-time electricity market. *Energies* **2018**, *11*, 2658.
8. Kath, C.; Ziel, F. The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. *Energy Economics* **2018**, *76*, 411–423.
9. Maciejowska, K.; Nitka, W.; Weron, T. Day-ahead vs. Intraday – Forecasting the price spread to maximize economic benefits. *Energies* **2019**, *12*, 631.
10. Uniejewski, B.; Marcjasz, G.; Weron, R. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. *International Journal of Forecasting* **2019**, *35*, 1533–1547.
11. Janke, T.; Steinke, F. Forecasting the price distribution of continuous intraday electricity trading. *Energies* **2019**, *12*, 4262.
12. Narajewski, M.; Ziel, F. Estimation and simulation of the transaction arrival process in intraday electricity markets. *Energies* **2019**, *12*, 4518.
13. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*; Springer, New York, 2013.
14. 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.
15. Giacomini, R.; White, H. Tests of conditional predictive ability. *Econometrica* **2006**, *74*, 1545–1578.
16. Uniejewski, B.; Weron, R.; Ziel, F. Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems* **2018**, *33*, 2219–2229.
17. Uniejewski, B.; Weron, R. Efficient forecasting of electricity spot prices with expert and LASSO models. *Energies* **2018**, *11*, 2039.
18. Ziel, F.; Weron, R. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics* **2018**, *70*, 396–420.
19. Uniejewski, B.; Nowotarski, J.; Weron, R. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* **2016**, *9*, 621.
20. Monteiro, C.; Ramirez-Rosado, I.; Fernandez-Jimenez, L.; Conde, P. Short-term price forecasting models based on artificial neural networks for intraday sessions in the Iberian electricity market. *Energies* **2016**, *9*, 721.
21. Andrade, J.; Filipe, J.; Reis, M.; Bessa, R. Probabilistic price forecasting for day-ahead and intraday markets: Beyond the statistical model. *Sustainability* **2017**, *9*, 1990.
22. Ziel, F. Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems* **2016**, *31*, 4977–4987.
23. Elliott, G.; Timmermann, A. *Economic Forecasting*; Princeton University Press, 2016.
24. Diebold, F.X.; Shin, M. Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. *International Journal of Forecasting* **2018**, *35*, 1679–1691.
25. 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.
26. 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.
27. Nowotarski, J.; Weron, R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews* **2018**, *81*, 1548–1568.
28. Muniain, P.; Ziel, F. Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. *International Journal of Forecasting* **2020**. (DOI: 10.1016/j.ijforecast.2019.11.006).
29. Oksuz, I.; Ugurlu, U. Neural network based model comparison for intraday electricity price forecasting. *Energies* **2019**, *12*, 4557.

30. Hubicka, K.; Marcjasz, G.; Weron, R. A note on averaging day-ahead electricity price forecasts across calibration windows. *IEEE Transactions on Sustainable Energy* **2019**, *10*, 321–323.