

Efficient Forecasting of Electricity Spot Prices with Expert and LASSO Models

Authors:

Bartosz Uniejewski, Rafał Weron

Date Submitted: 2018-09-21

Keywords: variance stabilizing transformation, automated variable selection, LASSO, long-term seasonal component, day-ahead market, electricity spot price

Abstract:

Recent electricity price forecasting (EPF) studies suggest that the least absolute shrinkage and selection operator (LASSO) leads to well performing models that are generally better than those obtained from other variable selection schemes. By conducting an empirical study involving datasets from two major power markets (Nord Pool and PJM Interconnection), three expert models, two multi-parameter regression (called baseline) models and four variance stabilizing transformations combined with the seasonal component approach, we discuss the optimal way of implementing the LASSO. We show that using a complex baseline model with nearly 400 explanatory variables, a well chosen variance stabilizing transformation (asinh or N-PIT), and a procedure that recalibrates the LASSO regularization parameter once or twice a day indeed leads to significant accuracy gains compared to the typically considered EPF models. Moreover, by analyzing the structures of the best LASSO-estimated models, we identify the most important explanatory variables and thus provide guidelines to structuring better performing models.

Record Type: Published Article

Submitted To: LAPSE (Living Archive for Process Systems Engineering)

Citation (overall record, always the latest version):

LAPSE:2018.0561

Citation (this specific file, latest version):

LAPSE:2018.0561-1

Citation (this specific file, this version):


LAPSE:2018.0561-1v1

DOI of Published Version: <https://doi.org/10.3390/en11082039>

License: Creative Commons Attribution 4.0 International (CC BY 4.0)

Article

Efficient Forecasting of Electricity Spot Prices with Expert and LASSO Models

Bartosz Uniejewski ^{1,2} and Rafał Weron ^{1,*} 

¹ Department of Operations Research, Faculty of Computer Science and Management, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; uniejewskibartosz@gmail.com

² Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

* Correspondence: rafal.weron@pwr.edu.pl; Tel.: +48-71-320-4525

Received: 29 June 2018; Accepted: 30 July 2018; Published: 6 August 2018



Abstract: Recent *electricity price forecasting* (EPF) studies suggest that the *least absolute shrinkage and selection operator* (LASSO) leads to well performing models that are generally better than those obtained from other variable selection schemes. By conducting an empirical study involving datasets from two major power markets (Nord Pool and PJM Interconnection), three expert models, two multi-parameter regression (called *baseline*) models and four variance stabilizing transformations combined with the *seasonal component* approach, we discuss the optimal way of implementing the LASSO. We show that using a complex baseline model with nearly 400 explanatory variables, a well chosen variance stabilizing transformation (asinh or N-PIT), and a procedure that recalibrates the LASSO regularization parameter once or twice a day indeed leads to significant accuracy gains compared to the typically considered EPF models. Moreover, by analyzing the structures of the best LASSO-estimated models, we identify the most important explanatory variables and thus provide guidelines to structuring better performing models.

Keywords: electricity spot price; day-ahead market; long-term seasonal component; LASSO; automated variable selection; variance stabilizing transformation

1. Introduction

One of the challenges of short-term *electricity price forecasting* (EPF) is variable (or feature) selection. However, there is no standard approach when the number of potential explanatory variables is large. So far, the typical approach has been to select predictors in an ad-hoc fashion or by using expert knowledge [1]. Rarely has a formal validation procedure been carried out, and hardly ever has an automated selection or shrinkage algorithm been utilized.

The earliest known examples of statistically sound variable selection in EPF include Karakatsani and Bunn [2] and Misiorek [3] who used *stepwise regression* to eliminate statistically insignificant variables in parsimonious autoregression (AR) and regime-switching models in a multivariate setting, i.e., separately for the individual load periods; see [4] for a discussion of the uni- and multivariate frameworks in EPF. In the machine learning literature, Amjady and Keynia [5] proposed a feature selection algorithm based on *mutual information* that was later utilized in [6,7], among others. On the other hand, Gianfreda and Grossi [8] used a very simple technique—single-step elimination of insignificant predictors in a regression setting.

To our best knowledge, Barnes and Balda [9] were the first to apply regularization in EPF. Namely, in a study concerning the profitability of battery storage, they utilized *ridge regression* to compute forecasts of New York Independent System Operator (NYISO) electricity prices for a model with more than 50 regressors. By ‘putting big data analytics to work’—as they referred to it—Ludwig et al. [10]

used *random forests* and the *least absolute shrinkage and selection operator* (LASSO or Lasso) as a feature selection algorithm to choose which of the 77 available weather stations were relevant. In parallel, in the machine learning literature, González et al. [11] utilized random forests, while Keles et al. [7] combined the *k-Nearest-Neighbor* algorithm with *backward elimination* to select the most appropriate inputs out of more than 50 fundamental parameters or lagged versions of these parameters.

The next qualitative change came with the papers by Ziel et al. [12,13], who used LASSO to sparsify very large (100+) sets of model parameters, either utilizing B-splines in a univariate setting or, more efficiently, within a multivariate framework. In the first thorough comparative study, Uniejewski et al. [14] evaluated six automated selection and shrinkage procedures (single-step elimination, forward and backward stepwise regression, ridge regression, LASSO, and *elastic nets*) applied to a baseline model with 100+ regressors. They concluded that using LASSO and elastic nets can bring significant accuracy gains compared to commonly used EPF models. Finally, in a study on the optimal model structure for short-term EPF, i.e., uni- vs. multivariate, Ziel and Weron [4] considered autoregressive models with 200+ potential explanatory variables (but not exogenous) and concluded that both uni- and multivariate LASSO-implied structures significantly outperform autoregressive benchmarks and that combining their forecasts can bring further improvements in predictive accuracy.

Yet, despite these very recent advances in the use of regularization techniques for EPF—and particularly the LASSO—a few issues remain open:

- **Optimal structure of the baseline model.** Which of the available regressors should we consider? Should the pool of regressors be parsimonious, limited mostly to lagged prices as in references [4,12–14]? Or the contrary, should we include as many regressors as possible, in particular, lagged exogenous variables?
- **The choice of the LASSO tuning (or regularization) parameter λ .** So far, two schemes have been proposed in the EPF literature. Ziel et al. [4,12,13] selected one λ for each day and hour in the test period by maximizing an information criterion in the 730-day in-sample calibration window, while Uniejewski et al. [14] used one 91-day cross-validation window to select one λ for all days and hours in the test period. Which approach leads to more accurate forecasts—frequent recalibration (every day and hour), one λ for all days and hours, or something in between? What is the optimal window length for selecting λ ?
- **The use of variance stabilizing transformations (VSTs).** Recent studies advocate utilizing the *seasonal component* approach of Nowotarski and Weron [15] to eliminate long-term trend-seasonal effects and applying a VST to ‘normalize’ the marginal price distributions prior to fitting a forecasting model. So far, only parsimonious regression and neural net structures have been tested. However, is this approach also beneficial in case of parameter-rich and LASSO-estimated models? If yes, which of the VSTs discussed and evaluated in Uniejewski et al. [16] are to be recommended?

To this end, we conduct a comprehensive empirical study involving nearly 5-year long datasets from two major power markets (Nord Pool and PJM Interconnection), multiple model structures (including three expert models, two multi-parameter regression *baseline* models and LASSO-estimated models based on the latter two structures), a range of alternative procedures for estimating λ , four variance stabilizing transformations combined with the *seasonal component* approach, and an evaluation in terms of the classical error measure for point forecasts (i.e., the *mean absolute error*, MAE) and the Diebold–Mariano (DM) test [17] to determine significant differences in forecasting accuracy.

For the benchmarks, we consider a **naïve** similar-day scheme and three parsimonious autoregressive models with exogenous variables (ARX). Two of them are built on popular structures, originally proposed by Misiorek et al. [18] and Ziel and Weron [4], and the third one is a new specification that was motivated by the results presented in Table 2 of Uniejewski et al. [14]. Following the later paper and Ziel [13], we refer to all three ARX benchmarks as *expert* models, since they are built on some prior knowledge of experts. As the baseline structures that are ‘shrunk’ via LASSO,

we use two multi-parameter ARX models—one has been already considered in the EPF context [14], and the other is a newly proposed specification. Furthermore, we propose a new two-step procedure that merges automated variable selection via LASSO with ordinary least squares (OLS) estimation of its weights and compare the obtained point forecasts across four variance stabilizing transformations (VSTs) and 39 schemes of selecting the tuning parameter (λ).

The remainder of the paper is structured as follows. In Section 2, we first introduce the two datasets and the rolling window scheme, and then discuss seasonal decomposition and the four VSTs used in this study. In Section 3, we introduce the considered model structures. Then, in Section 4, we compare the predictive performance in terms of MAE and the DM test across models, VSTs, and λ -selection schemes. We also analyze the structures of the best LASSO-estimated models to identify the most important explanatory variables and thus provide guidelines to structuring better performing models. Finally, in Section 5, we wrap up the results and conclude.

2. Methodology

2.1. The Test Ground

For the test ground we chose two major power markets—Nord Pool (Europe) and PJM (United States). The first dataset comprises system prices, day-ahead *consumption prognosis* for four Nordic countries (Denmark, Finland, Norway, and Sweden), and day-ahead wind power *generation prognosis* for the period 1 January 2013 to 19 September 2017, all at hourly resolution. The time series plotted in Figure 1 were constructed using data published by the Nordic power exchange Nord Pool (www.nordpoolspot.com) and preprocessed to account for missing values and changes to/from daylight saving time, analogously as in [19]. The missing data values (corresponding to the changes to the daylight saving/summer time and eight hourly consumption figures for Norway) were substituted by the arithmetic average of the neighboring values. The ‘doubled’ values (corresponding to the changes from the daylight saving/summer time) were substituted by the arithmetic average of the two values for the ‘doubled’ hour.

The second dataset comprises hourly prices for the Commonwealth Edison (COMED) zone in the PJM market and day-ahead system (RTO) and zonal (COMED) load forecasts for the same period, i.e., 1 January 2013 to 19 September 2017 (see Figure 2). The time series were constructed using data published by the PJM (www.pjm.com) and preprocessed to account for missing values and changes to or from the daylight saving time, analogously, as for the Nord Pool dataset. Note, that although the electricity price units are similar in Figures 1 and 2 (EUR/MWh vs. USD/MWh), the scales on the y-axis are different because of the much more volatile behavior of the PJM market, particularly in 2014.

Like the majority of EPF studies, we considered a rolling window scheme. To avoid seasonal artifacts, we let the calibration window length be a multiple of the weekly and annual periodicities. On the other hand, due to a large number of variables considered (as many as 386 in one of the models), we needed a window of more than 386 days to avoid an underdetermined system with fewer equations than unknowns. Hence, the choice of a circa two-year calibration window of 728 days or 104 full weeks which is the shortest window that satisfies these criteria.

Each day, the 728-day long calibration window was rolled forward by 24 h and price forecasts for the 24 h of the next day were computed. Initially, all considered models (their short-term and long-term components) were calibrated with data from 1 January 2013 to 19 September 2017 and forecasts for all 24 h of 30 December 2014 were determined. Then, the window was rolled forward by one day and forecasts for all 24 h of the next day are computed. This procedure is repeated until predictions for the last day in the sample (i.e., 19 September 2017) were made.

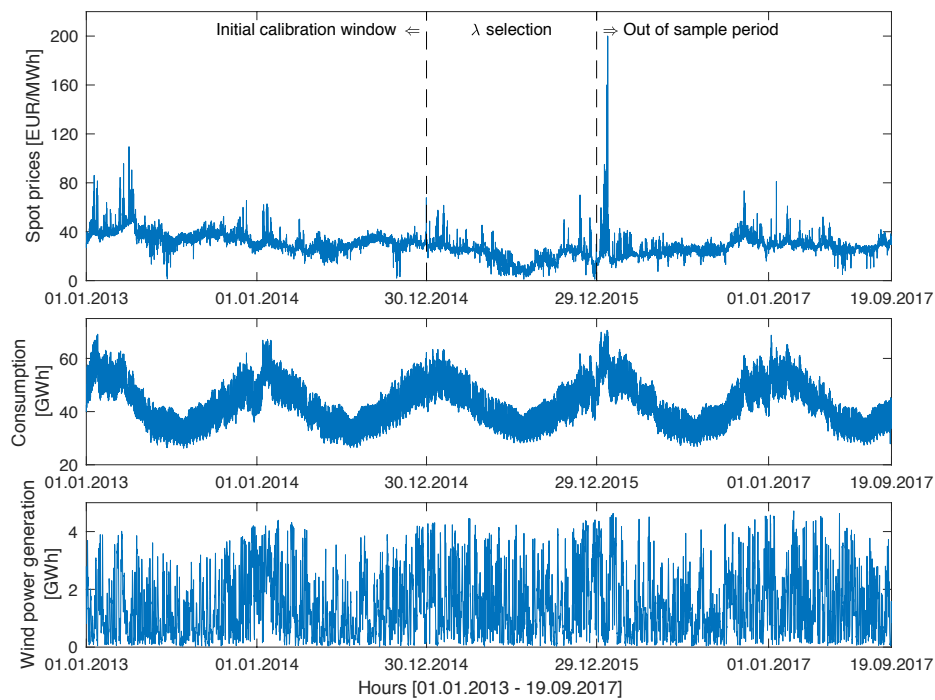


Figure 1. Nord Pool hourly system prices (top), hourly consumption prognosis (middle) and hourly wind power prognosis for Denmark (bottom) for the period 1 January 2013 to 19 September 2017. The vertical dashed lines mark the beginning of the 364-day period for selecting λ s for LASSO and the beginning of the 631-day long out-of-sample test period.

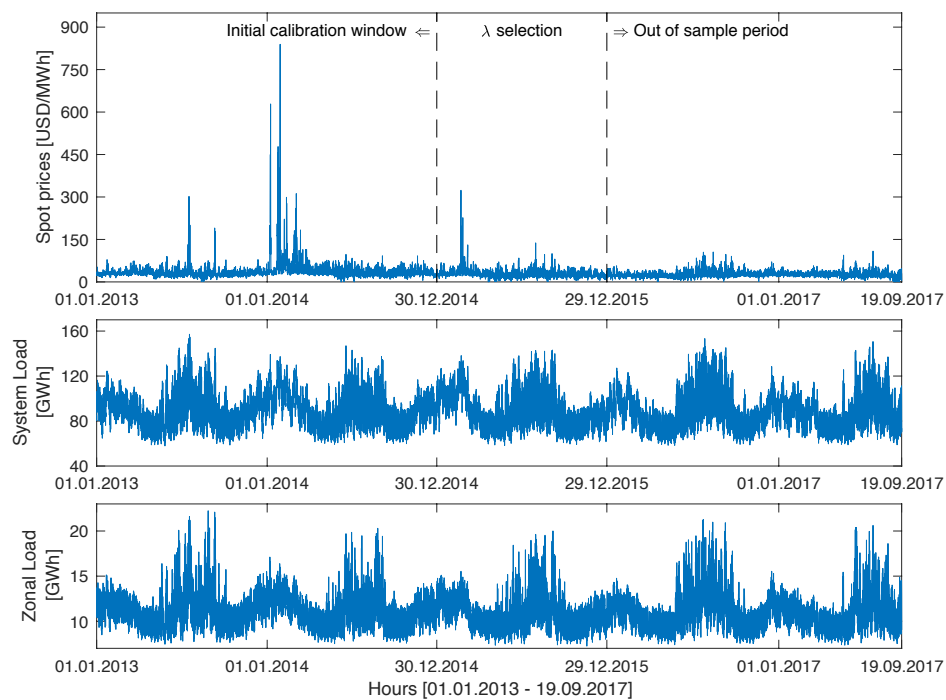


Figure 2. Hourly prices for the Commonwealth Edison (COMED) zone in the PJM market (top), hourly system load forecasts (middle) and hourly zonal load forecasts for the COMED zone (bottom) for the period 1 January 2013 to 19 September 2017. The dashed lines have the same meaning as in Figure 1.

2.2. Seasonal Decomposition

Recent published research leaves no doubt that seasonal decomposition of the price series leads to significant improvement in forecasting accuracy [15,20,21]. Here, we followed the *Seasonal Component Autoregressive* (SCAR) modeling framework of Nowotarski and Weron [15], i.e., the series of electricity spot prices is decomposed in an additive manner into a *long-term seasonal component* (LTSC) and the remaining stochastic component with short-term periodicities. Then, the two were modeled separately and their day-ahead forecasts are added up to yield the final price forecast. Note that following Uniejewski et al. [21], we added an additional step to the original SCAR algorithm, in which we decomposed the exogenous series (consumption prognosis and wind power generation prognosis; see Section 2.1).

Moreover, in contrast to other papers built around the *seasonal component* concept, (i) we did not limit the analysis to log-transformed data but considered more general variance stabilizing transformations (see Section 2.3) and (ii) worked only with the S_9 wavelet-based LTSC [19,22] which has been found to be robust and well performing across a range of datasets and forecasting models [20,21].

2.3. Variance Stabilizing Transformations

The basic idea behind a *variance stabilizing transformation* (VST) is to reduce the variation in price data. Lower variation and/or less spiky behavior of the input data usually allows the forecasting model to yield more accurate predictions [22]. In our models, we let $Y_{d,h}$ denote the deseasonalized electricity price in the day-ahead market for day d and hour h , with $X_{d,h}$ being transformed data, i.e., $X_{d,h} = f(Y_{d,h})$, where $f(\cdot)$ is a given VST. After computing the forecasts, we applied the inverse transformation to obtain the deseasonalized electricity spot price forecasts: $\hat{Y}_{d,h} = f^{-1}(\hat{X}_{d,h})$. We used four VST classes that have been found to perform well in an EPF context [16]. Three of them work on ‘normalized’ values, $y_{d,h} = \frac{1}{b}(Y_{d,h} - a)$, where a is the median of $Y_{d,h}$ in the 728-day calibration window and b is the sample *median absolute deviation* (MAD) around the median. The last class builds on the *probability integral transform* (PIT) and does not require normalization. In Figure 3, we illustrate all four VSTs on a sample dataset.

The *area hyperbolic sine* transformation is nothing else but the inverse of the *hyperbolic sine*,

$$X_{d,h} = \mathbf{asinh}(y_{d,h}) \equiv \log \left(y_{d,h} + \sqrt{y_{d,h}^2 + 1} \right), \quad (1)$$

with inverse $y_{d,h} = \sinh(X_{d,h})$. It has been used in the context of electricity prices, e.g., by Schneider [23] and Ziel and Weron [4]. The *mirror-log* and *polynomial* transformations were introduced by Uniejewski et al. [16] as generalizations of the robust Box–Cox transformation for $\lambda = 0$ and $\lambda \neq 0$, respectively. The former is a straightforward generalization of the log-transformation with a mirror image of the logarithm for negative values:

$$X_{d,h} = \mathbf{mlog}(y_{d,h}, c) = \operatorname{sgn}(y_{d,h}) \left[\log \left(|y_{d,h}| + \frac{1}{c} \right) + \log(c) \right], \quad (2)$$

with inverse $y_{d,h} = \operatorname{sgn}(X_{d,h}) \left[e^{|X_{d,h}| - \log c} - \frac{1}{c} \right]$. It is a one parameter family. Here, we used $c = \frac{1}{3}$, as suggested by Uniejewski et al. [16]. Similarly to the standard log-transformation and **asinh**, the **mlog** exhibits a log-damping effect. The used *polynomial* transformation was

$$X_{d,h} = \mathbf{poly}(y_{d,h}, c, \lambda) = \operatorname{sgn}(y_{d,h}) \left[\left| |y_{d,h}| + \left(\frac{c}{\lambda} \right)^{\frac{1}{\lambda-1}} \right|^\lambda - \left(\frac{c}{\lambda} \right)^{\frac{\lambda}{\lambda-1}} \right], \quad (3)$$

with inverse $y_{d,h} = \operatorname{sgn}(X_{d,h}) \left[\left\{ |X_{d,h}| + \left(\frac{c}{\lambda} \right)^{\frac{\lambda}{\lambda-1}} \right\}^{\frac{1}{\lambda}} - \left(\frac{c}{\lambda} \right)^{\frac{1}{\lambda-1}} \right]$. It is a two parameter family. Here, we used $\lambda = 0.125$ and $c = 0.05$, as suggested by Uniejewski et al. [16]. Both the **mlog** and **poly** had slopes of c at the origin.

The last class we considered is based on the so-called *probability integral transform*: $\text{PIT}(Z) = F_Z(Z)$, where F_Z is the cumulative distribution function (cdf) of Z . The **N-PIT** is defined as

$$X_{d,h} = \mathbf{N}\text{-PIT}(Y_{d,h}) \equiv \Phi^{-1}(\text{PIT}(Y_{d,h})) = \Phi^{-1}(\hat{F}_Y(Y_{d,h})), \quad (4)$$

with the inverse $Y_{d,h} = \hat{F}_Y^{-1}(\Phi(X_{d,h}))$, where Φ is the standard normal cdf. The **N-PIT**, misleadingly called the Nataf transformation, was used by Diaz and Planas [24] to normalize Spanish electricity prices. It was also used together with the t-PIT (a Student-t transformed PIT) in a recent extensive EPF study [16]. In Figure 3, we can clearly see that the histogram of **N-PIT**-transformed prices looks like the standard normal density.

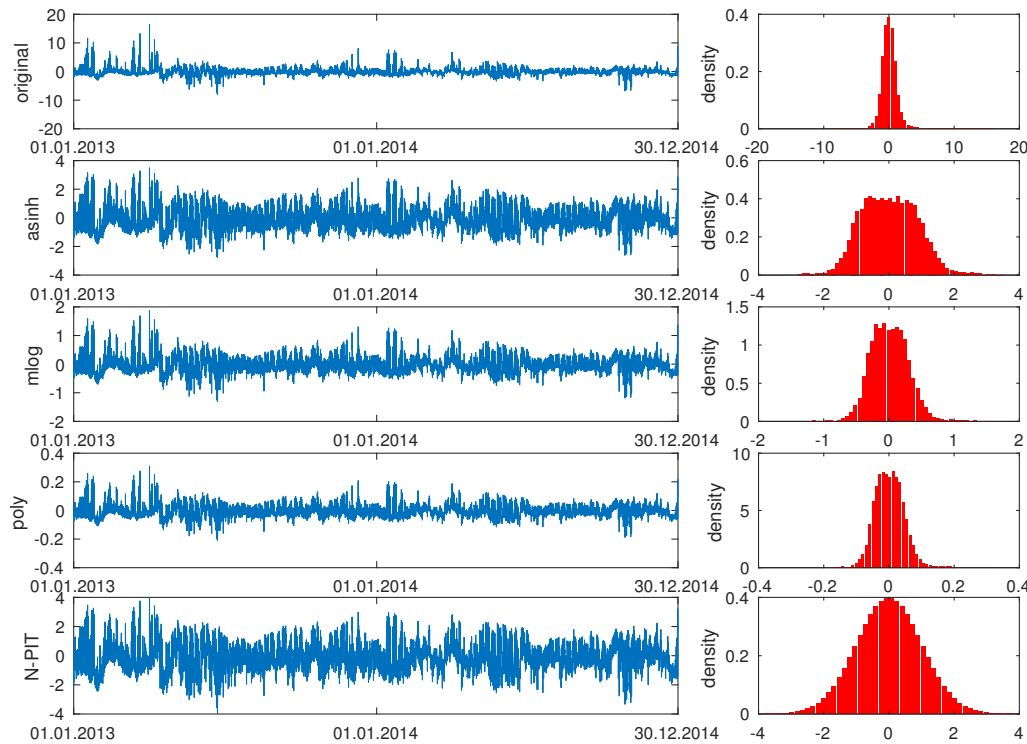


Figure 3. Time series plots of the original (in EUR/MWh), **asinh**-transformed, **mlog**-transformed, **poly**-transformed and **N-PIT**-transformed Nord Pool spot prices (from top to bottom). All series were deseasonalized with respect to the S_9 wavelet-based long-term seasonal component (LTSC) beforehand. The resulting marginal densities are depicted in the right panels.

2.4. The Forecasting Framework

Let us now summarize the forecasting framework. For each autoregressive model discussed in Section 3, i.e., all models except the **naive** benchmark, this consisted of five steps:

1. (a) Decompose the series of prices in the calibration window of $728 \times 24 = 17,472$ h using S_9 wavelet smoothing, i.e., utilizing nine decomposition levels [19,22], into a trend-seasonal and a stochastic component (with short-term periodicities): $P_{d,h} = T_{d,h} + Y_{d,h}$.
- (b) Compute persistent forecasts of the LTSC independently for each of the 24 h of the next day, i.e., $\hat{T}_{d^*+1,h} \equiv T_{d^*,h}$ for $h = 1, \dots, 24$, where d^* is the last day in the calibration window.
- (c) Decompose the exogenous series in the calibration window of $(728 + 1) \times 24 = 17,496$ h using the same type of wavelet smoothing as in Step 1(a). Note that we can start one day earlier since the load and wind power forecasts are known one day in advance.
2. Normalize $y_{d,h} = \frac{1}{b}(Y_{d,h} - a)$ and transform $X_{d,h} = f(y_{d,h})$, the stochastic component, where f is one of the VSTs defined in Section 2.3 and $(a, b) = (\text{median}, \text{MAD})$ of $Y_{d,h}$ in the

calibration window. Analogously normalize and transform the stochastic component(s) of the exogenous series.

3. Calibrate one of the autoregressive models discussed in Section 3 to $X_{d,h}$ and compute forecasts for each hour of the next day: $\widehat{X}_{d^*+1,h}, h = 1, \dots, 24$.
4. Back-transform the obtained forecasts: $\widehat{Y}_{d^*+1,h} = bf^{-1}(\widehat{X}_{d^*+1,h}) + a$.
5. Add the persistent LTSC forecasts to yield the final price forecasts: $\widehat{P}_{d^*+1,h} = \widehat{T}_{d^*+1,h} + \widehat{Y}_{d^*+1,h}$.

Note that compared to recent EPF papers utilizing the seasonal component approach [15,20,21], we introduced important changes. Firstly, instead of the logarithm we used one of four VSTs defined in Section 2.3, which can also be applied to price series with negative values, like those from the German European Energy Exchange (EEX) [25]. Secondly, the order of data transformation and decomposition was swapped. We first seasonally decomposed the data, and then transformed the stochastic part; the same applied for the exogenous series (consumption and wind power generation prognosis).

3. Models

3.1. Benchmarks

The first benchmark, the so-called **naive** method, belongs to the class of similar-day techniques. It is defined by $\widehat{P}_{d,h} = P_{d-7,h}$ for Monday, Saturday and Sunday, and $\widehat{P}_{d,h} = P_{d-1,h}$ otherwise. As Contreras et al. [26] argued, forecasting procedures that are not calibrated carefully fail to outperform this extremely simple method surprisingly often.

The second benchmark, denoted by **ARX1**, is a simple autoregressive structure originally proposed by Misiorek et al. [18] and later used in a number of EPF studies [13,14,19,20,27–30]:

$$X_{d,h} = \beta_{h,1}X_{d-1,h} + \beta_{h,2}X_{d-2,h} + \beta_{h,3}X_{d-7,h} + \beta_{h,4}X_{d-1}^{min} + \beta_{h,5}C_{d,h} + \beta_{h,6}D_{Sat} + \beta_{h,7}D_{Sun} + \beta_{h,8}D_{Mon} + \varepsilon_{d,h}, \quad (5)$$

where $X_{d-1,h}$, $X_{d-2,h}$ and $X_{d-7,h}$ refer to the transformed and deseasonalized price series. They account for the autoregressive effects of the previous days (the same hour yesterday, two days ago, and one week ago), while X_{d-1}^{min} is the minimum of the previous day's 24 hourly preprocessed prices. The exogenous variable $C_{d,h}$ refers to the prognosis of consumption for day d and hour h (actually, to a transformed and deseasonalized series). The three dummy variables D_{Sat} , D_{Sun} and D_{Mon} account for the weekly seasonality, and are defined as $D_i = 1$ for $d = i$ and zero otherwise. Finally, the $\varepsilon_{d,h}$ s are assumed to be independent and identically distributed normal variables.

The third benchmark was the *expert_{DOW,nl}* model of Ziel and Weron [4], only expanded to include one exogenous variable (consumption prognosis). The model, denoted by **ARX2**, is given by the following formula:

$$X_{d,h} = \beta_{h,1}X_{d-1,h} + \beta_{h,2}X_{d-2,h} + \beta_{h,3}X_{d-7,h} + \beta_{h,4}X_{d-1,24} + \beta_{h,5}X_{d-1}^{min} + \beta_{h,6}X_{d-1}^{max} + \beta_{h,7}C_{d,h} + \sum_{i=1}^7 \beta_{h,8,i}D_i + \varepsilon_{d,h}. \quad (6)$$

The price for the last load period of the previous day, i.e., $X_{d-1,24}$, is included to take advantage of the fact that prices for early morning hours depend more on the previous day's price at midnight than on the price for the same hour [13,31]. Compared to **ARX1**, the maximum of the previous day's prices, i.e., X_{d-1}^{max} , and dummies for the remaining days of the week ($D_4 \equiv D_{Tue}$, $D_5 \equiv D_{Wed}$, $D_6 \equiv D_{Thu}$, $D_7 \equiv D_{Fri}$) are also included. As there are seven dummies in Equation (6), one of them plays the role of the intercept. Note, that there is no intercept in Equation (5); this is not a major shortcoming since the prices are deseasonalized (i.e., centered) beforehand.

The last benchmark, denoted by **ARX3**, is based on the results of Uniejewski et al. [14]. The authors used shrinkage methods (particularly, LASSO and elastic nets) to find the most commonly selected regressors of a multi-variable ARX model. We analyzed these results and used them to construct a benchmark model that combines the knowledge of experts and the outcome of automated variable selection. The model is defined by

$$\begin{aligned} X_{d,h} = & \beta_{h,1}X_{d-1,h} + \beta_{h,2}X_{d-7,h} + \beta_{h,3}X_{d-1,h-1} + \beta_{h,4}X_{d-1,h+1} \\ & + \beta_{h,5}X_{d-1,21} + \beta_{h,6}X_{d-1,22} + \beta_{h,7}X_{d-1,24} + \beta_{h,8}X_{d-2,24} \\ & + \beta_{h,9}C_{d,h} + \beta_{h,10}C_{d-7,h} + \beta_{h,11}W_{d,h} + \sum_{i=1}^7 \beta_{h,12,i}D_i + \varepsilon_{d,h}. \end{aligned} \quad (7)$$

Compared to **ARX2**, this benchmark includes more autoregressive terms and two additional exogenous variables: consumption prognosis for the previous week, i.e., $C_{d-7,h}$, and wind power generation prognosis (for Nord Pool) or zonal load forecast (for PJM) of the forecasted day and hour, i.e., $W_{d,h}$. Actually, the latter two refer to transformed and deseasonalized values.

3.2. Baseline Autoregressive Models

An indisputable advantage of using automatic variable selection is an almost unlimited number of initially considered explanatory variables. As a result, the knowledge of experts becomes less important. However, the human factor is not completely eliminated. The forecaster still has to choose an appropriate baseline model that includes all variables of (potential) interest. To see how much this choice impacts the forecasting accuracy, we considered two different baseline models. The first one, denoted by **bARX1**, is essentially the **fARX** model of Uniejewski et al. [14] with consumption prognosis $C_{d,h}$ as the first exogenous variable and wind power generation prognosis (for Nord Pool) or zonal load forecast (for PJM) $W_{d,h}$ as the second, both after decomposition and transformation:

$$\begin{aligned} X_{d,h} = & \sum_{i=1}^3 \sum_{j=1}^{24} \beta_{h,i,j,1} X_{d-i,j} + \beta_{h,7,h,1} X_{d-7,h} \\ & + \sum_{i=1}^3 \left(\beta_{h,i,1,2} X_{d-i}^{min} + \beta_{h,i,1,3} X_{d-i}^{max} \right) \\ & + \beta_{h,0,h,4} C_{d,h} + \beta_{h,1,h,4} C_{d-1,h} + \beta_{h,7,h,4} C_{d-7,h} + \beta_{h,0,h,5} W_{d,j} \\ & + \sum_{i=1}^7 \beta_{h,i,1,6} D_i + \sum_{i=1}^7 \beta_{h,i,1,7} D_i X_{d-1,h} + \sum_{i=1}^7 \beta_{h,i,1,8} D_i C_{d,h} + \varepsilon_{d,h}. \end{aligned} \quad (8)$$

Note that in contrast to the benchmark autoregressive models discussed in Section 3.1, we treated holidays as the eighth day of the week, i.e., $D_1 = \dots = D_7 = 0$ for national holidays in Norway (for Nord Pool) or in the United States (for PJM); they were identified using the Time and Date AS web page: www.timeanddate.com/holidays. Consequently, like the **ARX1** model, but unlike **ARX2** and **ARX3**, the regression in Equation (8) has no intercept.

The second baseline model, denoted by **bARX2**, is much richer (386 vs. 104 variables). It uses information about prices from all hours of the whole previous week, the exogenous variables play much more important roles (175 vs. 4 variables) and the part accounting for weekly seasonality is more complex (the dummies are additionally multiplied by the average price from the previous day and the last known spot price, transformed, and deseasonalized). The model is given by

$$\begin{aligned}
X_{d,h} = & \sum_{i=1}^7 \sum_{j=1}^{24} \beta_{h,i,j,1} X_{d-i,j} + \sum_{i=1}^7 \left(\beta_{h,i,1,2} X_{d-i}^{\min} + \beta_{h,i,1,3} X_{d-i}^{\max} \right) \\
& + \sum_{i=\{0,1,2,7\}} \sum_{j=1}^{24} \beta_{h,i,j,4} C_{d-i,j} + \sum_{i=\{0,1,7\}} \sum_{j=1}^{24} \beta_{h,i,j,5} W_{d-i,j} \\
& + \sum_{i=1}^8 \beta_{h,i,1,6} D_i + \sum_{i=1}^7 \beta_{h,i,1,7} D_i X_{d-1,h} + \sum_{i=1}^7 \beta_{h,i,1,8} D_i X_{d-1,24} \\
& + \sum_{i=1}^7 \beta_{h,i,1,9} D_i X_{d-1}^{avg} + \sum_{i=1}^7 \beta_{h,i,1,10} D_i C_{d,h} + \varepsilon_{d,h}. \tag{9}
\end{aligned}$$

Note that for $h = 24$, $\beta_{h,i,1,8} \equiv 0$ for all i . Moreover, compared to **BARX1**, the model includes an explicit holiday dummy variable (D_8); hence, the regression in Equation (9) includes an intercept.

3.3. LASSO-Estimated Models

Let us write the structure of both baseline models in a more compact form:

$$X_{d,h} = \sum_{i=1}^n \beta_{h,i} V_{d,h}^i \tag{10}$$

where the $V_{d,h}^i$ s are the regressors, and the $\beta_{h,i}$ s are the corresponding coefficients. *Shrinkage* (also known as *regularization*) fits the full model with all predictors using an algorithm that shrinks the coefficients of the less important explanatory variables towards zero [32]. The *least absolute shrinkage and selection operator* (LASSO or Lasso) of Tibshirani [33] may actually shrink some of the coefficients to zero and hence perform variable selection.

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:

$$\hat{\beta}_h^L = \min_{\beta} \{ \text{RSS} + \lambda \|\beta_h\|_1 \} = \min_{\beta} \left\{ \text{RSS} + \lambda \sum_{i=1}^n |\beta_{h,i}| \right\}, \tag{11}$$

where $\lambda \geq 0$ is a *tuning* (or *regularization*) parameter. Note that for $\lambda = 0$, we get the standard least squares estimator, for $\lambda \rightarrow \infty$, all $\beta_{h,i}$ s tend to zero, while for intermediate values of λ , there is a balance between minimizing the RSS and shrinking the coefficients towards zero (and each other). Selecting a ‘good’ value for λ is critical. However, to our best knowledge, the optimal choice of λ has not been discussed in the EPF context so far. We compared alternative λ -selection schemes for a grid of 25 exponentially decreasing λ s ranging from 10^0 to 10^{-6} , and selected the tuning parameter for which the mean absolute (prediction) error in the D -day validation period was the smallest.

Models for which the LASSO operator is used to estimate parameters are denoted by $*\text{Lasso}_i^D$, where $i = 1, 2$ refers to one of two baseline models (respectively **BARX1** and **BARX2**); D is the length of the validation window; and the asterisk represents the λ -selection scheme. We considered eight window lengths— $D = 1$ (one day), 7 (one week), 28 (one ‘month’), 60 (two months), 91 (one quarter), 182 (half a year), 273 (three quarters) and 364 (one year)—and six λ -selection schemes:

- $24 \times \text{NLasso}_i^D$, where for each day and each hour in the test period, one λ is selected,
- $2 \times \text{NLasso}_i^D$, where for each day in the test period, two λ s are selected—one for the on-peak and one for the off-peak hours,
- $1 \times \text{NLasso}_i^D$, where for each day in the test period, one λ is selected,
- 24Lasso_i^D , where for each hour, one λ is selected for the whole test period,

- **2Lasso** $_i^D$, where two λ s are selected (on-peak and off-peak) for the whole test period,
- **1Lasso** $_i^D$, where one λ is selected for the whole test period.

3.4. LassOLS-Type Models

To answer the question of how important the shrinkage property is for reducing the prediction errors, we introduced a modification of the ***Lasso** $_i^D$ class of models, where the LASSO operator is used to select the explanatory variables but not to estimate the β_h s; instead, we used OLS to obtain parameter estimates for the LASSO-selected variables. In this way, we automatically created 24 different ‘expert’, LASSO-implied models for each day. Since these model structures are dependent on the choice of λ , like in Section 3.3, we considered six model schemes: $nxN\text{LassOLS}_i^D$ and $n\text{LassOLS}_i^D$, with $n = 1, 2, 24$, which correspond to $nxN\text{Lasso}_i^D$ and $n\text{Lasso}_i^D$ models, respectively.

4. Empirical Results

4.1. Evaluation in Terms of MAE and the DM Test

As the main evaluation criterion, we considered the *mean absolute error* (MAE) for the full out-of-sample test period of $D = 631$ days, i.e., 29 December 2015 to 19 September 2017:

$$\text{MAE} = \|\mathcal{E}_{d,h}\|_1 = \frac{1}{24D} \sum_{d=1}^D \sum_{h=1}^{24} |\mathcal{E}_{d,h}|, \quad (12)$$

where $\mathcal{E}_{d,h} = P_{d,h} - \hat{P}_{d,h}$ is the prediction error for a given day d and hour h . Additionally, we used the Diebold–Mariano (DM) test [17] to draw statistically significant conclusions on the outperformance of the forecasts of one model by those of another. In the EPF literature, the DM test is usually performed separately for each hour of the day [1]. However, Ziel and Weron [4] recently introduced a different approach, where only one statistic for each pair of models is computed based on the 24-dimensional vector of errors for each day:

$$\Delta_{X,Y,d} = \|\mathcal{E}_{X,d}\| - \|\mathcal{E}_{Y,d}\|, \quad (13)$$

where $\|\mathcal{E}_{Z,d}\| = \sum_{h=1}^{24} |\mathcal{E}_{d,h}|$ for model Z . For each model pair and each dataset, we computed the p -value of two one-sided DM tests: (i) a test with the null hypothesis $H_0 : E(\Delta_{X,Y,d}) \leq 0$, i.e., the outperformance of the forecasts of Y by those of X , and (ii) the complementary test with the reverse null $H_0^R : E(\Delta_{X,Y,d}) \geq 0$, i.e., the outperformance of the forecasts of X by those of Y . As in the standard DM test, we assumed that the loss differential series was covariance stationary.

4.2. Performance across Model Classes and VSTs

In Table 1, we report the MAEs in the 631-day test period for 30 considered models and five VSTs (Id denotes the identity ‘transformation’). The models include four benchmarks, two baseline models, 12 ***LassOLS** $_i^{91}$ models and 12 ***Lasso** $_i^{91}$ models, i.e., LASSO-type models with a 91-day validation window to select λ . In Figure 4 and 5, the results of the ‘multivariate’ DM-test for the same 30 models and the **N-PIT** transformation (i.e., for both datasets corresponding to the last column in Table 1) are summarized, while Figure 6 shows the two best performers in Table 1 across the VSTs (i.e., corresponding to rows 26–27 in Table 1). The ‘chessboards’ present the p -values of the conducted pairwise comparisons in graphical form. The green and yellow squares indicate statistical significance at the 5% level, with the darkest green corresponding to close to zero p -values. The red squares indicate weak significance with p -values between 5% and 10%, while black squares denote no significance, i.e., p -values of 10% or more. Several interesting conclusions can be drawn.

Table 1. Mean absolute errors (MAE) in the 631-day test period for 30 considered models (in rows), five variance stabilizing transformations (VSTs; in columns; the first ‘transformation’ is identity, **Id**) and two markets (Nord Pool and PJM). The models include four benchmarks, two baseline models, 12 ***LassOLS**_{*i*}⁹¹ models and 12 ***Lasso**_{*i*}⁹¹ models. Note that all LASSO-type models use a 91-day validation window for selecting λ . A heat map is used to indicate better (→green) and worse (→red) results.

MAE	Nord Pool					PJM				
	Id	asinh	mlog	poly	NPIT	Id	asinh	mlog	poly	NPIT
naive	2.3338					3.8531				
ARX1	1.9254	1.8491	1.8467	1.8482	1.8143	2.9246	2.8172	2.8270	2.8303	2.7847
ARX2	1.7506	1.6422	1.6263	1.6272	1.6363	2.9032	2.6986	2.7174	2.7240	2.6668
ARX3	1.6713	1.4831	1.4773	1.4805	1.4910	2.8457	2.6369	2.6609	2.6681	2.6110
bARX1	1.8020	1.5447	1.5413	1.5518	1.5519	2.8327	2.4504	2.4968	2.5119	2.4241
bARX2	2.5270	2.1671	2.1400	2.1362	2.0143	3.8500	3.1544	3.2424	3.2665	3.2476
24xNLassOLS ₁	1.6368	1.4486	1.4570	1.4602	1.4563	2.6868	2.3976	2.4319	2.4489	2.3453
2xNLassOLS ₁	1.6609	1.4536	1.4561	1.4635	1.4484	2.7209	2.4172	2.4477	2.4616	2.3654
1xNLassOLS ₁	1.6709	1.4480	1.4611	1.4720	1.4525	2.7345	2.4324	2.4525	2.4643	2.3770
24LassOLS ₁	1.6719	1.4530	1.4603	1.4698	1.4569	3.3577	2.5241	2.6683	2.8116	2.4075
2LassOLS ₁	1.6847	1.4540	1.4623	1.4771	1.4551	3.4289	2.4473	2.6712	2.5928	2.4130
1LassOLS ₁	1.6855	1.4540	1.4578	1.4706	1.4551	4.1461	2.4422	2.6712	2.5440	2.4341
24xNLasso ₁	1.5786	1.4356	1.4321	1.4363	1.4522	2.5974	2.3589	2.3838	2.4024	2.3281
2xNLasso ₁	1.5976	1.4334	1.4338	1.4379	1.4516	2.6092	2.3638	2.3924	2.4082	2.3421
1xNLasso ₁	1.6187	1.4337	1.4315	1.4415	1.4519	2.6258	2.3717	2.4039	2.4174	2.3391
24Lasso ₁	1.6178	1.4377	1.4331	1.4419	1.4558	3.3907	2.4495	2.5710	2.6056	2.3820
2Lasso ₁	1.6179	1.4349	1.4299	1.4470	1.4524	3.1617	2.4096	2.5423	2.5735	2.3680
1Lasso ₁	1.6207	1.4416	1.4378	1.4436	1.4524	3.1617	2.4096	2.5423	2.6335	2.3508
24xNLassOLS ₂	1.6641	1.3799	1.3800	1.3936	1.3640	2.6875	2.4027	2.4275	2.4570	2.3627
2xNLassOLS ₂	1.7021	1.3801	1.3821	1.3954	1.3656	2.7116	2.4089	2.4411	2.4581	2.3676
1xNLassOLS ₂	1.6778	1.3749	1.3818	1.3942	1.3604	2.7465	2.3985	2.4607	2.4676	2.3864
24LassOLS ₂	1.6802	1.3829	1.3906	1.4056	1.3852	3.4260	2.4477	2.6310	2.9983	2.4850
2LassOLS ₂	1.6759	1.3713	1.3846	1.4084	1.3715	3.0317	2.3888	2.9658	2.7753	2.3662
1LassOLS ₂	1.6818	1.3725	1.3794	1.4084	1.3715	3.4588	2.3877	2.6618	2.8809	2.3686
24xNLasso ₂	1.5766	1.3483	1.3800	1.3936	1.3454	2.5703	2.3239	2.3487	2.3648	2.3211
2xNLasso ₂	1.5849	1.3401	1.3821	1.3954	1.3404	2.5712	2.3144	2.3393	2.3598	2.3070
1xNLasso ₂	1.5917	1.3391	1.3818	1.3942	1.3427	2.5873	2.3332	2.3530	2.3698	2.3053
24Lasso ₂	1.5996	1.3459	1.3906	1.4056	1.3587	3.3100	2.3607	2.5252	2.5524	2.3704
2Lasso ₂	1.6031	1.3419	1.3846	1.4084	1.3565	2.9388	2.3245	2.4429	2.5294	2.3060
1Lasso ₂	1.5999	1.3419	1.3794	1.4084	1.3565	2.9388	2.3139	2.4890	2.4646	2.3060

- All considered models (except **bARX2** with the identity transform, i.e., without applying a VST) beat the **naive** benchmark by a large margin, see Table 1 and Figure 4. This indicates that they all are highly efficient forecasting tools. The poor performance of **bARX2** may be due to identification issues of a model with nearly 400 explanatory variables calibrated with as few as 728 daily observations.
- On the other hand, the smaller baseline model (**bARX1**) performed reasonably well, especially for PJM data, where it significantly outperformed other benchmarks. This is in contrast to the results reported in Uniejewski et al. [14] for the Nord Pool market. The reason may be that for a regression model with 104 variables, the increased length of the calibration window (from one to two years) leads to a much better performance.
- LASSO-type models based on **bARX2** significantly outperformed all others (the lower left ‘quadrant’ of the chessboards in Figures 4 and 5 is black and the upper right ‘quadrant’ is green). The latter is more visible in Figure 4 for the Nord Pool market. In particular, for Nord Pool data, LASSO-type models improved the accuracy by ca. 7% in comparison to the same models based on the smaller **bARX1** baseline model.

- The LASSO operator treated as a shrinkage technique (i.e., a forecasting tool on its own; ***Lasso**_{*i*}^{*D*} models) significantly outperformed the two-step procedure where LASSO is only used for variable selection and OLS is used for estimation (***LassOLS**_{*i*}^{*D*} models).
- The **ARX3** model strongly outperformed the more parsimonious benchmarks, obtaining for Nord Pool data results that are almost as good as those from LASSO-based models built on **bARX1**(see Table 1 and the relatively black row in Figure 4 corresponding to the **ARX3** model).
- The choice of the best performing transformation depends on the complexity of the model, e.g., for LASSO models based on **bARX2**, the **asinh** and **N-PIT** transformations led to the most accurate forecasts (see Table 1), and the benefit from using a VST increased with the complexity of the model. Moreover, for the two best performing models, a significant difference between using the **asinh** and **N-PIT** appeared only once (for PJM data) and it was in favor of the **N-PIT** transformation. Note, that both VSTs led to significantly better predictions than the remaining ones (see Figure 6).
- For almost all cases, models that re-estimate (reselect) λ on a daily basis, i.e., $n \times \text{NLassOLS}_i^D$ and $n \times \text{NLasso}_i^D$, outperformed models with a fixed tuning parameter, i.e., $n \text{LassOLS}_i^D$ and $n \text{Lasso}_i^D$, see Table 1 and Figure 4. On the other hand, re-estimating λ is ca. 25 times slower (assuming that we have a grid of 25 λ s as we do here). The computational time can be reduced, however, by utilizing the results presented in Figure 7, i.e., by limiting the grid to 3–6 values of λ for $D > 60$ days.

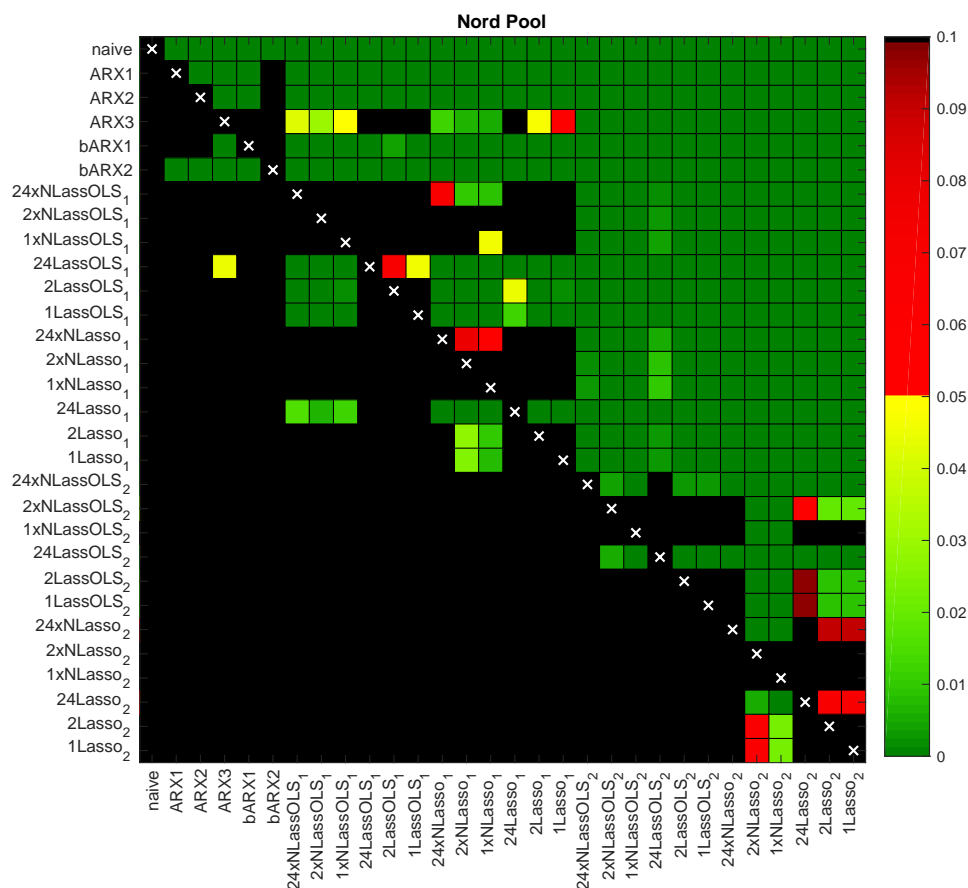


Figure 4. Results of the ‘multivariate’ variance stabilizing transformations (VSTs) multi-parameter Diebold–Mariano (DM) test for Nord Pool for the same 30 models as in Table 1 and the **N-PIT** transformation. A heat map is used to indicate the range of the *p*-values—the closer they are to zero (→dark green), the more significant the difference is between the forecasts of a model on the X-axis (better) and the forecasts of a model on the Y-axis (worse).

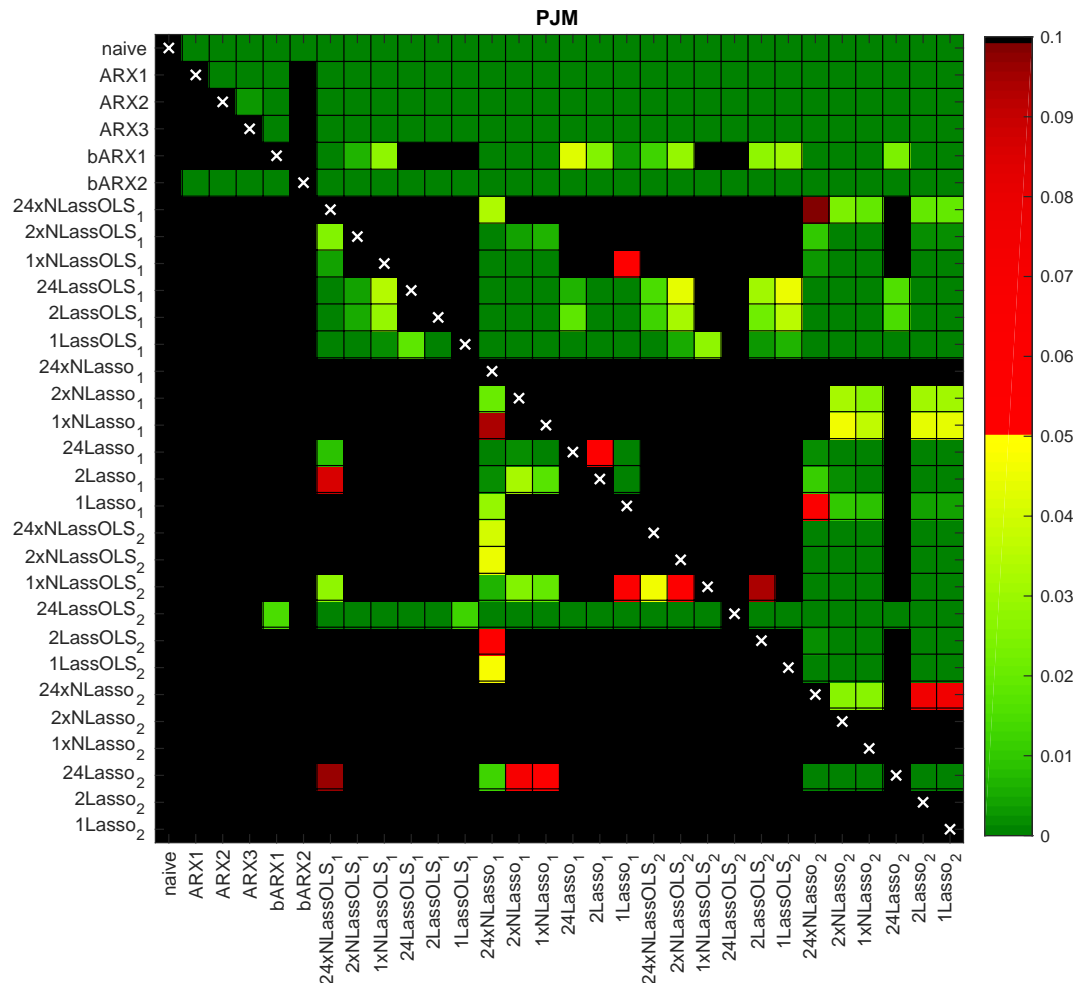


Figure 5. Results of the ‘multivariate’ DM test for PJM for the same 30 models as in Table 1 and the N-PIT transformation. A heat map is used to indicate the range of the p -values—the closer they are to zero (\rightarrow dark green), the more significant the difference is between the forecasts of a model on the X-axis (better) and the forecasts of a model on the Y-axis (worse).

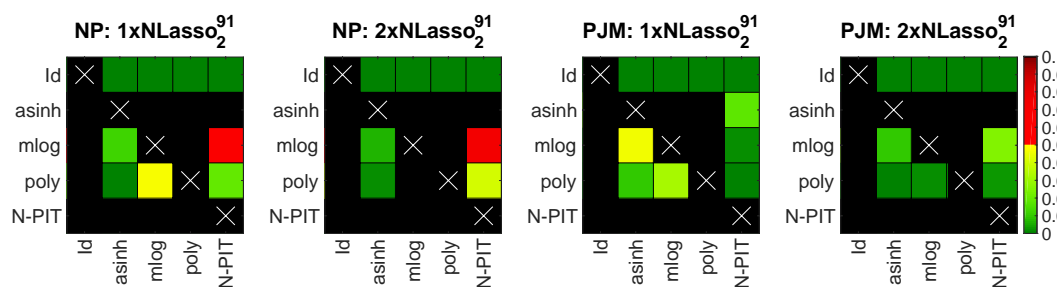


Figure 6. Results of the ‘multivariate’ DM test for the best performers in Table 1 (i.e., $1xNLasso_2^{91}$ and $2xNLasso_2^{91}$) for Nord Pool (two left panels) and PJM (two right panels) across the VSTs. Like in Figures 4 and 5, a heat map is used to indicate the range of obtained p -values for the pairwise comparisons.

To better understand the impact of applying a VST, it is necessary to refer to the article of Uniejewski et al. [16], where 16 transformations were tested across 12 datasets. Importantly, only models without exogenous variables were used in the cited study—an AR counterpart of our ARX2 and its neural network equivalent—and the price series were not decomposed into the LTSC and the stochastic component. The N-PIT turned out to be the best performing VST overall, at least in terms of MAE. However, for Nord Pool system prices covering a similar time period as in our

study (30 July 2010 to 28 July 2016), the **N-PIT** was worse than **asinh**, **mlog**, and **poly**. Note that the same conclusion holds here for the autoregressive benchmarks (including **ARX2**). This suggests that the conclusions of Uniejewski et al. [16] are likely to remain unchanged when using a model with exogenous variables and/or a model that is fitted to deseasonalized series. From this perspective, the **N-PIT** gains the most from model complexity and even for more complex models, the **N-PIT** would probably outperform other VSTs.

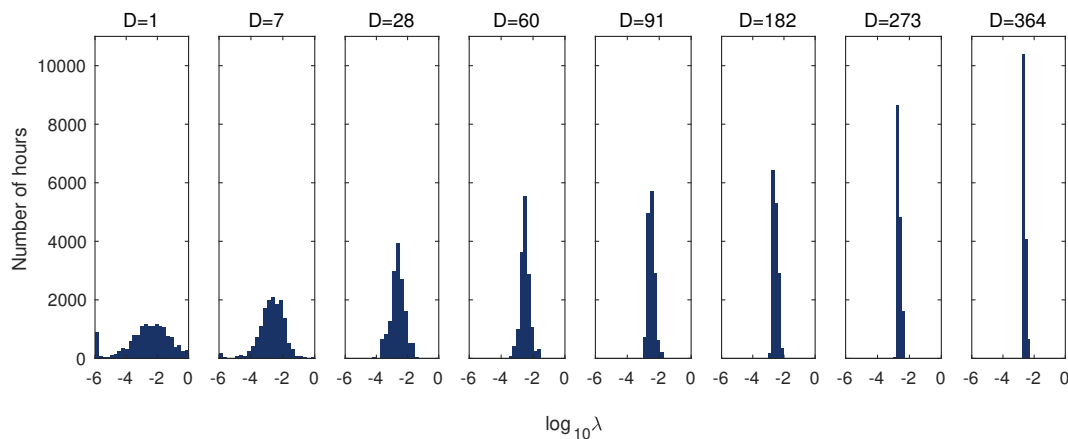


Figure 7. Histograms displaying the number of hours for which a given value of λ was selected in the whole Nord Pool test period ($24 \times 631 = 15,144$ h) across different D s and for the **N-PIT**-transformed $2xNLasso_2^D$ model. Recall, that we considered 25 exponentially decreasing λ s ranging from 10^0 to 10^{-6} ; they are plotted on a logarithmic scale on the X -axis. The histograms for PJM data look very similar (not plotted).

4.3. Performance across λ -Selection Schemes

Up to this point, all LASSO-type models have used a 91-day validation window for selecting λ . A pertinent question arises as to whether this is the optimal choice for D . To address this, we considered eight validation window lengths: $D = 1$ (one day), 7 (one week), 28 (one ‘month’), 60 (two months), 91 (one quarter), 182 (half a year), 273 (three quarters) and 364 (one year; this window is illustrated in Figure 1). On the other hand, the results discussed in Section 4.2 clearly show that (i) the **bARX2** baseline model leads to significantly better performing LASSO-type models than the smaller **bARX1** model, and (ii) $*Lasso_i^{91}$ models significantly outperform $*LassOLS_i^{91}$. Hence, to simplify the comparison, in this section, we limit the analysis to $*Lasso_2^D$ models.

In Table 2, we report the MAEs in the 631-day test period for 39 different λ -selection schemes and five VSTs (including the identity ‘transformation’). Note that, as opposed to Table 1, now, the coloring is independent for each column to better emphasize the differences between the D s. For both markets we can see the following:

- As observed in Section 4.2, models that re-estimate (reselect) λ on a daily basis, i.e., $nxNLasso_i^D$, outperform models with a fixed tuning parameter, i.e., $nLasso_i^D$. Yet, considering the significantly lower computational requirements, it is still reasonable to take the latter models as benchmarks.
- As the length of the λ -selection window increases, the extreme values of the tuning parameter are less often selected and finally, for $D = 364$ days, only a few intermediate λ s are used (for a sample illustration see Figure 7).
- Selecting validation windows of less than 60 days is not advisable. In the case of re-estimating (reselecting) λ s on a daily basis, i.e., in the $nxNLasso_i^D$ models, it is recommended to use much longer validation windows.
- We cannot definitely reject or accept the hypothesis that choosing λ for each hour separately leads to an improvement in the accuracy of the forecasts. Overall, the best results in terms of MAE were

obtained for the 2xNLasso₂ and 1xNLasso₂ models and $D \geq 273$ days, but the differences between them and other well performing models were insignificant (DM test values are not reported here).

Table 2. Mean absolute errors (MAE) in the 631-day test period for 39 different λ -selection schemes (in rows), five VSTs (in columns; the first ‘transformation’ is identity, **Id**) and two markets (Nord Pool and PJM). A heat map is used to indicate better (\rightarrow green) and worse (\rightarrow red) results. Note that this time, the coloring is independent for each column to better emphasize the differences between the D s.

		Nord Pool					PJM				
	D	Id	asinh	mlog	poly	N-PIT	Id	asinh	mlog	poly	N-PIT
24xNLasso ₂	1	1.8297	1.6629	1.6563	1.5979	1.5849	2.9111	2.7536	2.7666	2.7244	2.7232
	7	1.6757	1.4358	1.4713	1.4652	1.4169	2.6990	2.4784	2.4942	2.4953	2.4705
	28	1.5878	1.3752	1.4027	1.4130	1.3752	2.6036	2.3567	2.3853	2.4077	2.3717
	60	1.5750	1.3598	1.3844	1.3979	1.3666	2.5743	2.3231	2.3578	2.3668	2.3401
	91	1.5850	1.3571	1.3808	1.3935	1.3670	2.5703	2.3239	2.3487	2.3648	2.3211
	182	1.5777	1.3533	1.3831	1.3984	1.3490	2.5867	2.3238	2.3616	2.3733	2.3104
	273	1.5822	1.3536	1.3811	1.3932	1.3450	2.5941	2.3209	2.3603	2.3787	2.3032
364	1.5766	1.3483	1.3800	1.3936	1.3454	2.6013	2.3250	2.3659	2.3868	2.3041	
2xNLasso ₂	1	1.7506	1.5411	1.5339	1.5359	1.5230	2.8190	2.5898	2.6115	2.6072	2.5754
	7	1.6539	1.4058	1.4437	1.4547	1.3795	2.6646	2.4235	2.4277	2.4309	2.4111
	28	1.6013	1.3641	1.3874	1.4081	1.3549	2.5956	2.3292	2.3632	2.3729	2.3367
	60	1.5787	1.3499	1.3776	1.3846	1.3452	2.5653	2.3053	2.3381	2.3563	2.3056
	91	1.5932	1.3425	1.3801	1.3849	1.3450	2.5712	2.3144	2.3393	2.3598	2.3070
	182	1.5881	1.3460	1.3801	1.3938	1.3413	2.5814	2.3239	2.3616	2.3798	2.2992
	273	1.5878	1.3483	1.3790	1.3954	1.3424	2.5806	2.3224	2.3716	2.3839	2.2998
364	1.5849	1.3401	1.3821	1.3954	1.3404	2.5875	2.3143	2.3695	2.3806	2.2945	
1xNLasso ₂	1	1.7390	1.4898	1.5333	1.4965	1.5102	2.7737	2.5476	2.5714	2.5554	2.5427
	7	1.6579	1.3908	1.4203	1.4315	1.3835	2.6689	2.4078	2.4174	2.4318	2.4090
	28	1.6015	1.3489	1.3929	1.4078	1.3506	2.6153	2.3413	2.3793	2.3936	2.3315
	60	1.5943	1.3474	1.3836	1.3885	1.3441	2.5922	2.3252	2.3591	2.3771	2.3054
	91	1.5919	1.3470	1.3809	1.3904	1.3434	2.5873	2.3332	2.3530	2.3698	2.3053
	182	1.5914	1.3456	1.3798	1.3939	1.3408	2.5761	2.3169	2.3555	2.3734	2.2965
	273	1.5923	1.3387	1.3838	1.3935	1.3411	2.5853	2.3151	2.3486	2.3733	2.2952
364	1.5917	1.3391	1.3818	1.3942	1.3427	2.5890	2.3163	2.3535	2.3751	2.2944	
24Lasso ₂	60	1.6070	1.3492	1.3946	1.4016	1.3629	3.4300	2.3831	2.6272	2.6881	2.3966
	91	1.6119	1.3542	1.3955	1.4065	1.3624	3.3100	2.3607	2.5252	2.5524	2.3704
	182	1.6104	1.3612	1.3954	1.4080	1.3645	2.9197	2.3433	2.4188	2.5131	2.3387
	273	1.6047	1.3514	1.3929	1.4024	1.3592	3.0088	2.3555	2.4240	2.5478	2.3352
	364	1.5996	1.3459	1.3906	1.4056	1.3587	3.0071	2.3829	2.4442	2.5258	2.3608
2Lasso ₂	60	1.6122	1.3419	1.3846	1.4165	1.3565	3.3539	2.4136	2.8478	2.5294	2.3567
	91	1.6122	1.3541	1.3846	1.4165	1.3565	2.9388	2.3245	2.4429	2.5294	2.3060
	182	1.6122	1.3541	1.3846	1.4165	1.3565	2.8792	2.3245	2.3760	2.4312	2.3060
	273	1.6122	1.3419	1.3846	1.4084	1.3565	2.7906	2.3347	2.3916	2.4646	2.3060
	364	1.6031	1.3419	1.3846	1.4084	1.3565	2.6686	2.3347	2.3760	2.4245	2.3246
1Lasso ₂	60	1.5999	1.3419	1.4087	1.4084	1.3565	3.5375	2.3139	2.4890	2.4646	2.3473
	91	1.5999	1.3419	1.4087	1.4084	1.3565	2.9388	2.3139	2.4890	2.4646	2.3060
	182	1.6353	1.3419	1.4087	1.4084	1.3565	2.7906	2.3139	2.3916	2.4646	2.3060
	273	1.5999	1.3419	1.4087	1.4084	1.3565	2.7906	2.3347	2.3916	2.4646	2.3060
	364	1.5999	1.3419	1.3794	1.4084	1.3565	2.6686	2.3347	2.3526	2.3915	2.3060

4.4. Variable Selection

Let us now comment on variable selection by looking at the 1xNLasso₂¹⁸² model calibrated to N-PIT-transformed data (the results for other *Lasso₂^D, $D > 60$ models were qualitatively similar). In Table 3 (for Nord Pool) and Table 4 (for PJM), we present the mean occurrence (in %) of model parameters across the 631-day test periods. A heat map is used to indicate more (\rightarrow green) and less (\rightarrow red) commonly selected variables. Several interesting conclusions can be drawn.

Table 3. Cont.

Table with 25 columns (labeled 1 to 24) and rows grouped by 'd' (d, d-1, d-2, d-7) and 'h' (1 to 24). The table is titled 'Load forecast (C_d,h)' and contains numerical data cells. Some cells are highlighted in red and green.

Table 3. Cont.

		d	h	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24						
Wind power generation forecast ($W_{d,h}$)	d	1	80	41	28	0	0	1	0	0	4	0	0	0	0	23	20	35	37	71	37	98	90	85	83	36							
		d-1	1	70	20	29	34	83	100	59	3	13	9	14	16	17	19	25	31	58	54	33	55	68	81	86	97						
			d-7	1	45	1	4	5	17	49	55	59	75	97	94	87	68	51	38	55	82	64	49	57	56	47	38						
				Dummies	Mon	92	83	60	32	14	0	11	0	29	49	17	4	0	0	10	30	6	4	0	10	30	6	0	0	0			
					D * $X_{d-1,h}$	Mon	0	0	0	0	0	18	98	84	56	41	25	17	18	15	20	21	20	20	32	5	11	18	29	14			
						D * X^{avg}	Mon	61	59	52	50	29	0	36	0	0	1	12	3	0	1	12	3	0	5	6	5	0	2	15	18		
							D * $C_{d,h}$	Mon	2	71	93	91	83	34	0	24	34	13	1	0	0	0	0	0	0	0	3	6	2	0	0	8	19
								D * $X_{d-1,24}$	Mon	0	0	1	0	0	3	0	86	100	94	88	84	77	81	85	68	87	99	98	88	54	42		

day $d - 1$ in Tables 3 and 4). The diagonal was less visible around midday and for Nord Pool, it almost disappeared. The latter may be attributed to the explanatory power of the day-ahead wind prognosis for this market.

- Lagged prices for the more distant days ($d - 2$ to $d - 7$) were much less often selected, see Tables 3 and 4. Although they cannot be ignored completely, particularly for days $d - 2$ and $d - 7$ for Nord Pool and day $d - 7$ for PJM, it is hard to find any regular patterns that could be rationally (economically) justified.
- As observed by Maciejowska and Nowotarski [31] and Ziel [13], the price for hour 24 (i.e., $X_{d-1,24}$) is an important predictor. For on-peak hours, however, the evening prices (i.e., $X_{d-1,h}$ for $h = 19, 20, 21, 22$) seem to have more explanatory power. What is also surprising is that the morning prices (5 a.m. to 9 a.m.) on day $d - 7$ provided more information than the evening prices on that day.
- The maximum prices on the previous seven days included more information than the minimum prices and for the Nord Pool dataset they were selected quite often (see the green cells in the bottom of Table 3). On the other hand, the minimum prices for day $d - 1$ had high explanatory power for the early morning hours (1 a.m. to 7 a.m.; this effect was less pronounced for the PJM market).
- The exogenous variables (load and wind generation forecasts) for the past days were seldom selected. Only the forecasts for day d had high explanatory power and generally, only around the diagonal, i.e., for the same or neighboring hours.
- Finally, the dummies and the dummy-linked prices for the same hour and for midnight were important predictors (see Tables 3 and 4). On the other hand, the dummy-linked load forecasts were of mixed importance, and the dummy-linked price averages for the previous day were usually removed by the LASSO (except for Tuesdays and Wednesdays for Nord Pool).

5. Conclusions

In this paper we presented a comprehensive empirical study on the optimal way of implementing the LASSO in electricity price forecasting that addresses the three open issues outlined in the Introduction. Concerning variable (or feature) selection, i.e., the optimal structure of the baseline model, we identified the most important variables and thus provided guidelines to structure better performing expert models. In particular, we found that large sets of potential regressors are not a problem for the LASSO procedure. Although the LASSO typically uses only a small fraction of the initial set of explanatory variables, providing additional information in the underlying model significantly improves the accuracy of the obtained forecasts. Nevertheless, lagged exogenous variables (i.e., load and wind generation forecasts for the past days) are seldom selected and can be ignored when building models. Only the day-ahead predictions for the target day have high explanatory power and generally, only for the same or neighboring hours. On the other hand, although lagged prices for the more distant days are much less often selected, they cannot be ignored completely (particularly days $d - 2$ and $d - 7$). We also confirmed the importance of dummies and the dummy-linked prices for the same hour and for midnight. At the same time, the dummy-linked load forecasts turned out to be of mixed explanatory power, while the dummy-linked price averages for the previous day were shown to be redundant.

Secondly, regarding the choice of the LASSO tuning parameter, we found one λ for all days and hours in the test period (as in Uniejewski et al. [14]) to be an acceptable option, but this is recommended only if the computational time needs to be significantly reduced. To increase the forecast accuracy, it is better to reselect λ on a daily basis. However, too frequent reselection (i.e., on an hourly basis, as in references [4,12,13]) is significantly outperformed by schemes that select only one or two (on-peak, off-peak) λ s per day, i.e., models 1xNLasso₂ and 2xNLasso₂, respectively. We also observed that despite a higher computational burden, it is extremely important to select λ over periods of at least 60 days; however, cross-validation windows of half a year or more are recommended.

Thirdly, concerning the novel concept of variance stabilizing transformations (VSTs), we confirmed and reinforced the conclusions of Uniejewski et al. [16]. Namely, we showed that the VSTs not only increase the forecasting accuracy of LASSO-estimated models—like in the case of parsimonious regression and neural net models – but that the gains from using an appropriate VST increase with the complexity of the model. Lastly, using the Diebold–Mariano test, we found that for more complex models, it is advisable to use the **asinh** or the **N-PIT** transformation.

Author Contributions: Conceptualization (B.U., R.W.); Investigation (B.U.); Software (B.U.); Supervision (R.W.); Validation (B.U., R.W.); Writing—original draft (B.U.); Writing—review & editing (R.W.).

Funding: This work was partially supported by the National Science Center (NCN, Poland) through grant no. 2015/17/B/HS4/00334.

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

References

1. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. [[CrossRef](#)]
2. Karakatsani, N.; Bunn, D. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *Int. J. Forecast.* **2008**, *24*, 764–785. [[CrossRef](#)]
3. Misiorek, A. Short-term forecasting of electricity prices: Do we need a different model for each hour? *Med. Econom. Toepass.* **2008**, *16*, 8–13.
4. Ziel, F.; Weron, R. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Econ.* **2018**, *70*, 396–420. [[CrossRef](#)]
5. Amjady, N.; Keynia, F. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. *IEEE Trans. Power Syst.* **2009**, *24*, 306–318. [[CrossRef](#)]
6. Voronin, S.; Partanen, J. Price forecasting in the day-ahead energy market by an iterative method with separate normal price and price spike frameworks. *Energies* **2013**, *6*, 5897–5920. [[CrossRef](#)]
7. Keles, D.; Scelle, J.; Paraschiv, F.; Fichtner, W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Appl. Energy* **2016**, *162*, 218–230. [[CrossRef](#)]
8. Gianfreda, A.; Grossi, L. Forecasting Italian electricity zonal prices with exogenous variables. *Energy Econ.* **2012**, *34*, 2228–2239. [[CrossRef](#)]
9. Barnes, A.K.; Balda, J.C. Sizing and economic assessment of energy storage with real-time pricing and ancillary services. In Proceedings of the 2013 4th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG), Rogers, AR, USA, 8–11 July 2013. [[CrossRef](#)]
10. Ludwig, N.; Feuerriegel, S.; Neumann, D. Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *J. Decis. Syst.* **2015**, *24*, 19–36. [[CrossRef](#)]
11. González, C.; Mira-McWilliams, J.; Juárez, I. Important variable assessment and electricity price forecasting based on regression tree models: Classification and regression trees, bagging and random forests. *IET Gener. Transm. Distrib.* **2015**, *9*, 1120–1128. [[CrossRef](#)]
12. Ziel, F.; Steinert, R.; Husmann, S. Efficient modeling and forecasting of electricity spot prices. *Energy Econ.* **2015**, *47*, 89–111. [[CrossRef](#)]
13. Ziel, F. Forecasting Electricity Spot Prices Using LASSO: On Capturing the Autoregressive Intraday Structure. *IEEE Trans. Power Syst.* **2016**, *31*, 4977–4987. [[CrossRef](#)]
14. Uniejewski, B.; Nowotarski, J.; Weron, R. Automated Variable Selection and Shrinkage for Day-Ahead Electricity Price Forecasting. *Energies* **2016**, *9*, 621. [[CrossRef](#)]
15. Nowotarski, J.; Weron, R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting. *Energy Econ.* **2016**, *57*, 228–235. [[CrossRef](#)]
16. Uniejewski, B.; Weron, R.; Ziel, F. Variance Stabilizing Transformations for Electricity Spot Price Forecasting. *IEEE Trans. Power Syst.* **2018**, *33*, 2219–2229. [[CrossRef](#)]
17. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263.
18. Misiorek, A.; Trück, S.; Weron, R. Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. *Stud. Nonlinear Dyn. Econom.* **2006**, *10*, 1558–3708. [[CrossRef](#)]

19. Weron, R. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*; John Wiley & Sons: Chichester, UK, 2006.
20. 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. *Int. J. Forecast.* **2018**. [[CrossRef](#)]
21. 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 Econ.* **2018**. [[CrossRef](#)]
22. Janczura, J.; Trück, S.; Weron, R.; Wolff, R. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Econ.* **2013**, *38*, 96–110. [[CrossRef](#)]
23. Schneider, S. Power spot price models with negative prices. *J. Energy Mark.* **2011**, *4*, 77–102. [[CrossRef](#)]
24. Diaz, G.; Planas, E. A Note on the Normalization of Spanish Electricity Spot Prices. *IEEE Trans. Power Syst.* **2016**, *31*, 2499–2500. [[CrossRef](#)]
25. Hagfors, L.; Kamperud, H.; Paraschiv, F.; Prokopczuk, M.; Sator, A.; Westgaard, S. Prediction of extreme price occurrences in the German day-ahead electricity market. *Quant. Financ.* **2016**, *16*, 1929–1948. [[CrossRef](#)]
26. Contreras, J.; Espínola, R.; Nogales, F.; Conejo, A. ARIMA models to predict next-day electricity prices. *IEEE Trans. Power Syst.* **2003**, *18*, 1014–1020. [[CrossRef](#)]
27. Gaillard, P.; Goude, Y.; Nedellec, R. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *Int. J. Forecast.* **2016**, *32*, 1038–1050. [[CrossRef](#)]
28. Kristiansen, T. Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy* **2012**, *49*, 328–332. [[CrossRef](#)]
29. Maciejowska, K.; Nowotarski, J.; Weron, R. Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *Int. J. Forecast.* **2016**, *32*, 957–965. [[CrossRef](#)]
30. Nowotarski, J.; Raviv, E.; Trück, S.; Weron, R. An empirical comparison of alternate schemes for combining electricity spot price forecasts. *Energy Econ.* **2014**, *46*, 395–412. [[CrossRef](#)]
31. Maciejowska, K.; Nowotarski, J. A hybrid model for GEFCom2014 probabilistic electricity price forecasting. *Int. J. Forecast.* **2016**, *32*, 1051–1056. [[CrossRef](#)]
32. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*; Springer: New York, NY, USA, 2013.
33. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **1996**, *58*, 267–288.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).