

Article

Automated Variable Selection and Shrinkage for Day-Ahead Electricity Price Forecasting

Bartosz Uniejewski, Jakub Nowotarski and Rafał Weron *

Department of Operations Research, Wrocław University of Technology, 50-370 Wrocław, Poland; uniejewskibartosz@gmail.com (B.U.); jakub.nowotarski@pwr.edu.pl (J.N.)

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

Academic Editor: Javier Contreras

Received: 5 July 2016; Accepted: 29 July 2016; Published: 5 August 2016

Abstract: In day-ahead electricity price forecasting (EPF) variable selection is a crucial issue. Conducting an empirical study involving state-of-the-art parsimonious expert models as benchmarks, datasets from three major power markets and five classes of automated selection and shrinkage procedures (single-step elimination, stepwise regression, ridge regression, lasso and elastic nets), we show that using the latter two classes can bring significant accuracy gains compared to commonly-used EPF models. In particular, one of the elastic nets, a class that has not been considered in EPF before, stands out as the best performing model overall.

Keywords: electricity price forecasting; day-ahead market; autoregression; variable selection; stepwise regression; ridge regression; lasso; elastic net

JEL: C14, C22, C51, C53, Q47

1. Introduction

Alongside short-term load forecasting, short-term electricity price forecasting (EPF) has become a core process of an energy company's operational activities [1]. The reason is quite simple. A 1% improvement in the mean absolute percentage error (MAPE) in forecasting accuracy would result in about 0.1%–0.35% cost reductions from short-term EPF [2]. In dollar terms, this would translate into savings of ca. \$1.5 million per year for a typical medium-size utility with a 5-GW peak load [3].

As has been noted in a number of studies, be it statistical or computational intelligence, a key point in EPF is the appropriate choice of explanatory variables [1,4–11]. The typical approach has been to select predictors in an ad hoc fashion, sometimes using expert knowledge, seldom based on some formal validation procedures. Very rarely has an automated selection or shrinkage procedure been carried out in EPF, especially for a large set of initial explanatory variables.

Early examples of formal variable selection in EPF include Karakatsani and Bunn [12] and Misiolek [13], who used stepwise regression to eliminate statistically insignificant variables in parsimonious autoregression (AR) and regime-switching models for individual load periods. Amjadi and Keynia [4] proposed a feature selection algorithm that utilized the mutual information technique. (for later applications, see, e.g., [11,14,15]). In an econometric setup, Gianfreda and Grossi [5] computed *p*-values of the coefficients of a regression model with autoregressive fractionally integrated moving average disturbances (Reg-ARFIMA) and in one step eliminated all statistically-insignificant variables. In a study concerning the profitability of battery storage, Barnes and Balda [16] utilized ridge regression to compute forecasts of the New York Independent System Operator (NYISO) electricity prices for a model with more than 50 regressors.

More recently, González et al. [17] used random forests to identify important explanatory variables among the 22 considered. Ludwig et al. [7] used both random forests and the least absolute shrinkage

and selection operator (i.e., lasso or LASSO) as a feature selection algorithm to choose the relevant out of the 77 available weather stations. In a recent neural network study, Keles et al. [11] combined the *k-nearest-neighbor* algorithm with backward elimination to select the most appropriate input variables out of more than 50 fundamental parameters or lagged versions of these parameters. Finally, Ziel et al. [9,18] used the lasso to sparsify very large sets of model parameters (well over 100). They used time-varying coefficients to capture the intra-day dependency structure, either using B-splines and one large regression model for all hours of the day [9] or, more efficiently, using a set of 24 regression models for the 24 h of the day [18].

However, a thorough study involving state-of-the-art parsimonious expert models as benchmarks, data from diverse power markets and, most importantly, a set of different selection or shrinkage procedures is still missing in the literature. In particular, to our best knowledge, elastic nets have not been applied in the EPF context at all. It is exactly the aim of this paper to address these issues. We perform an empirical study that involves:

- nine variants of three parsimonious autoregressive model structures with exogenous variables (ARX): one originally proposed by Misiorek et al. [19] and later used in a number of EPF studies [13,18,20–27], one which evolved from it during the successful participation of TEAM POLAND in the Global Energy Forecasting Competition 2014 (GEFCom2014; see [28–30]) and an extension of the former, which creates a stronger link with yesterday's prices and additionally considers a second exogenous variable (zonal load or wind power),
- three two-year long, hourly resolution test periods from three distinct power markets (GEFCom2014, Nord Pool and the U.K.),
- nine variants of five classes of selection and shrinkage procedures: single-step elimination of insignificant predictors (without or with constraints), stepwise regression (with forward selection or backward elimination), ridge regression, lasso and three elastic nets (with $\alpha = 0.25, 0.5$ or 0.75),
- model validation in terms of the robust weekly-weighted mean absolute error (WMAE; see [1]) and the Diebold–Mariano (DM; see [31]) test

and draw statistically-significant conclusions of high practical value.

The remainder of the paper is structured as follows. In Section 2, we introduce the datasets. Next, in Section 3, we first discuss the iterative calibration and forecasting scheme, then describe the techniques considered for price forecasting: a simple naive benchmark, nine variants of three parsimonious ARX-type model structures and five classes of selection and shrinkage procedures. In Section 4, we summarize the empirical findings. Namely, we evaluate the quality of point forecasts in terms of WMAE errors, run the DM tests to formally assess the significance of differences in the forecasting performance and analyze variable selection for the best performing elastic net model. Finally, in Section 5 wrap up the results and conclude.

2. Datasets

The datasets used in this empirical study include three spot market time series. The first one comes from the Global Energy Forecasting Competition 2014 (GEFCom2014), the largest energy forecasting competition to date [28]. The dataset includes three time series at hourly resolution: locational marginal prices, day-ahead predictions of system loads and day-ahead predictions of zonal loads and covers the period 1 January 2011–14 December 2013; see Figure 1. The origin of the data has never been revealed by the organizers. The full dataset is now available as supplementary material accompanying [28] (Appendix A); however, during the competition, the information set was being extended on a weekly basis to prevent 'peeking' into the future. The dataset was preprocessed by the organizers and does not include any missing or doubled values.

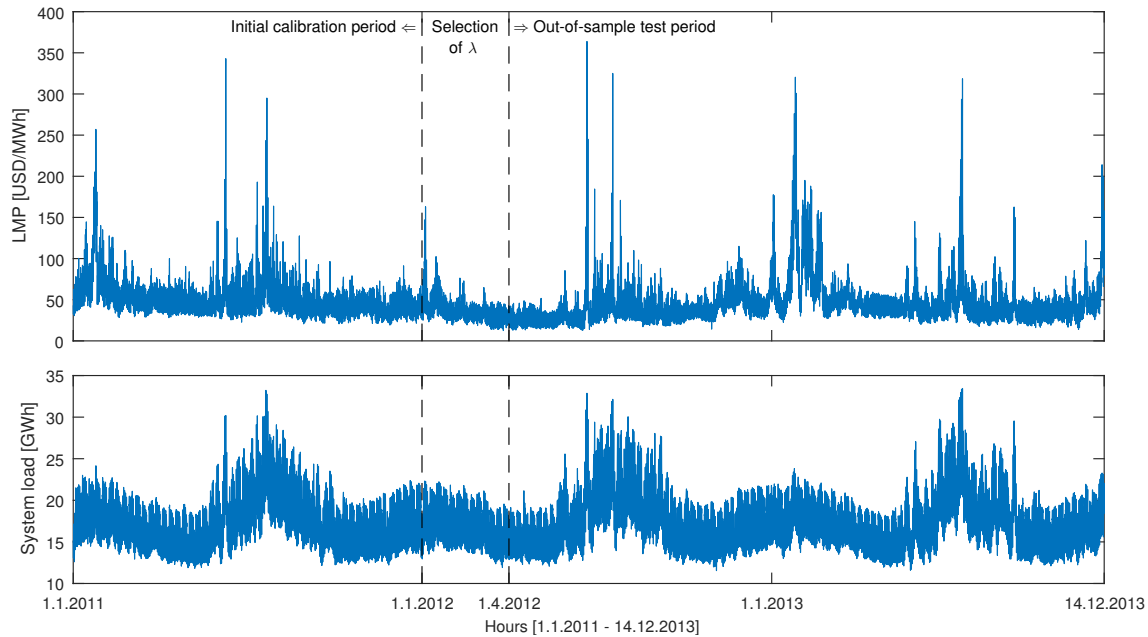


Figure 1. GEFCom2014 hourly locational marginal prices (LMP; top) and hourly day-ahead predictions of the system load (bottom) for the period 1 January 2011–14 December 2013. The day-ahead predictions of the zonal load are generally indistinguishable from those of the system load at this resolution; see Figure 8 in [28]. The vertical dashed lines mark the beginning of the 91-day period for selecting λ 's (ridge regression, lasso, elastic nets) and the beginning of the 623-day long out-of-sample test period.

The second dataset comes from one of the major European power markets: Nord Pool (NP). It comprises hourly system prices, hourly consumption prognosis for four Nordic countries (Denmark, Finland, Norway and Sweden) and hourly wind prognosis for Denmark and covers the period 1 January 2013–29 March 2016; see Figure 2. The time series 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 the daylight saving time, analogously as in [20] (Section 4.3.7). The missing data values (corresponding to the changes to the daylight saving/summer time; moreover, eight out of 28,392 hourly consumption figures were missing 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 third dataset comes from N2EX, the U.K. day-ahead power market operated by Nord Pool. It comprises hourly system prices for the period 1 January 2013–29 March 2016; see Figure 3. The time series was constructed using data published by Nord Pool (www.nordpoolspot.com) and, like the second dataset, preprocessed to account for changes to/from the daylight saving time. Note that the U.K. dataset includes only prices, as no day-ahead forecasts of fundamental variables were available to us. Hence, models calibrated to the U.K. data are 'pure price' models. To better see the effect of excluding fundamentals from forecasting models, we use the GEFCom2014 dataset twice, once with fundamentals (system and zonal load forecasts; to compare with the results for Nord Pool) and once without them.

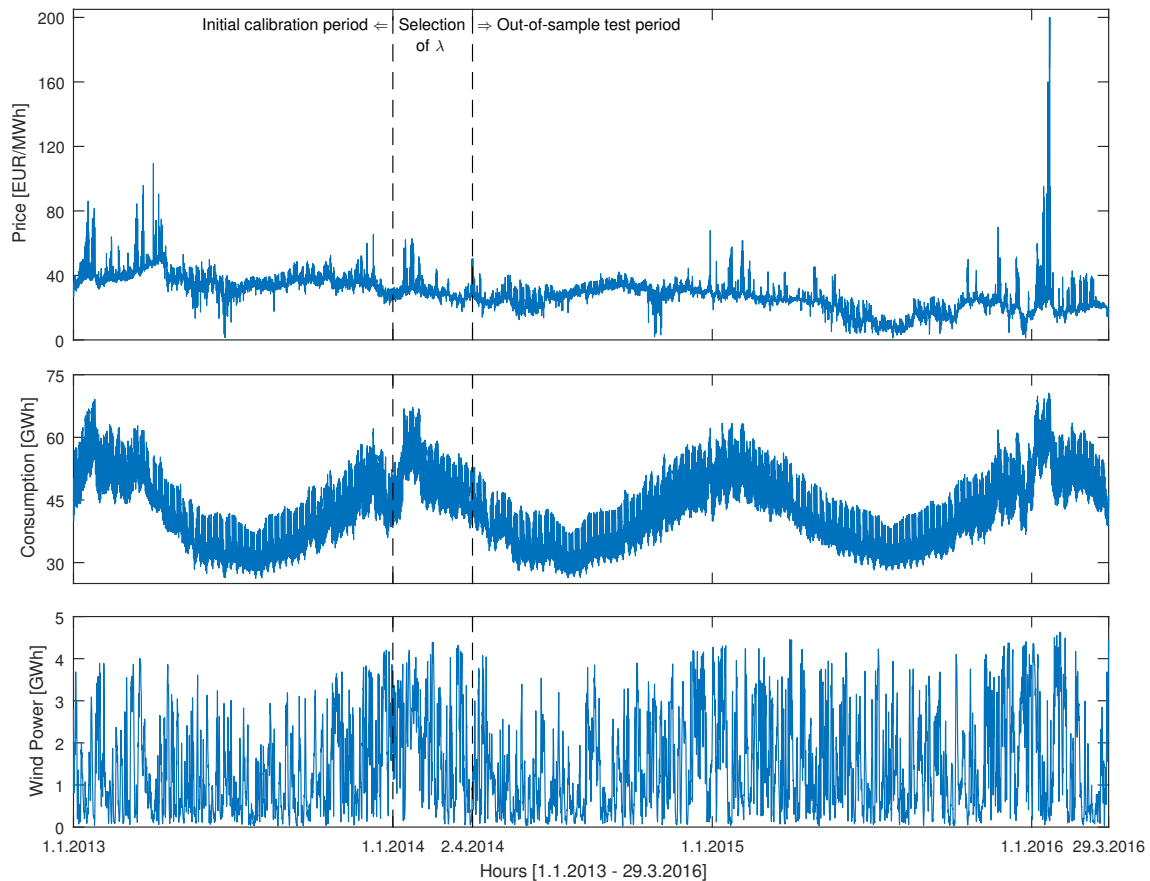


Figure 2. Nord Pool hourly system prices (top), hourly consumption prognosis (middle) and hourly wind power prognosis for Denmark (bottom) for the period 1 January 2013–29 March 2016. The vertical dashed lines mark the beginning of the 91-day period for selecting λ 's (ridge regression, lasso, elastic net) and the beginning of the 728-day long out-of-sample test period.

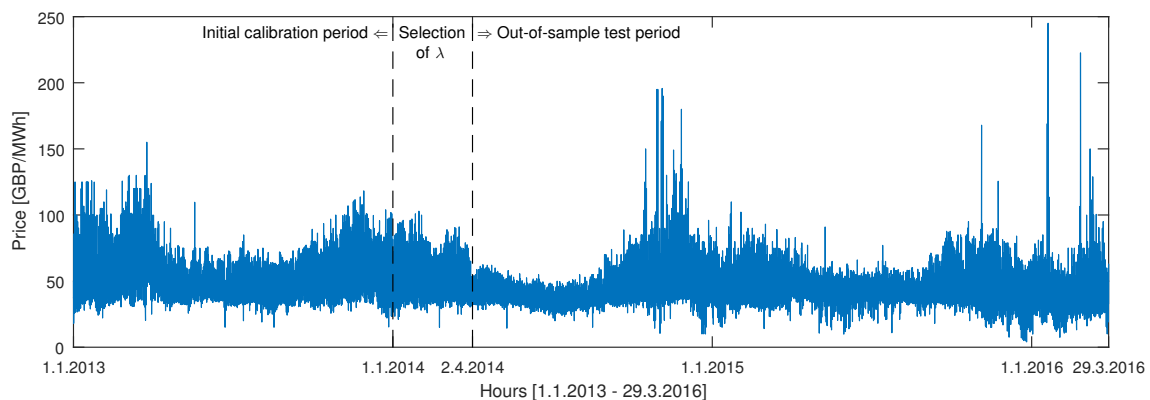


Figure 3. U.K. power market hourly system prices (Nord Pool's N2EX market) for the period 1 January 2013–29 March 2016. The vertical dashed lines mark the beginning of the 91-day period for selecting λ 's (ridge regression, lasso, elastic net) and the beginning of the 728-day long out-of-sample test period.

3. Methodology

It should be noted that although we use here the terms short-term, spot and day-ahead interchangeably, the former two do not necessarily refer to the day-ahead market. Short-term EPF generally involves predicting 24 hourly (or 48 half-hourly) prices in the day-ahead market, cleared

typically at noon on the day before delivery, i.e., 12–36 h before delivery, the adjustment markets, cleared a few hours before delivery, and the balancing or real-time markets, cleared minutes before delivery [32]. The spot market, especially in the literature on European electricity markets, is often used as a synonym of the day-ahead market. However, in the U.S., the spot market is another name for the real-time market, while the day-ahead market is called the forward market [20,33]. Furthermore, some markets in Europe nowadays admit continuous trading for individual load periods up to a few hours before delivery. With the shifting of volume from the day-ahead to intra-day markets, also in Europe, the term spot is more and more often being used to refer to the real-time markets [1].

Throughout this article, we denote by $P_{d,h}$ the electricity price in the day-ahead market for day d and hour h . Like many studies in the EPF literature [1], we use the logarithmic transform to make the price series more symmetric (see Figure 4) and compare with the top panels in Figures 1–3. We can do this since all considered datasets are positive-valued. However, this is not a very restrictive property. If datasets with zero or negative values were considered, we could work with non-transformed prices. Furthermore, we center the log-prices by subtracting their in-sample mean prior to parameter estimation. We do this independently for each hour $h = 1, \dots, 24$:

$$p_{d,h} = \log(P_{d,h}) - \frac{1}{T} \sum_{t=1}^T \log(P_{t,h}), \quad (1)$$

where T is the number of days in the calibration window; hence, the missing intercept ($\beta_{h,0} \equiv 0$) in our autoregressive models; for model parameterizations, see Sections 3.2–3.4.

For all three markets, the day-ahead forecasts of the hourly electricity price are determined within a rolling window scheme, using a 365-day calibration window. First, all considered models are calibrated to data from the initial calibration period (i.e., 1 January 2011–31 December 2011 for GEFCom2014 and 1 January 2013–31 December 2013 for Nord Pool and the U.K.), and forecasts for all 24 h of the next day (1 January) are determined. Then, the window is rolled forward by one day; the models are re-estimated, and forecasts for all 24 h of 2 January are computed. This procedure is repeated until the predictions for the 24 h of the last day in the sample (14 December 2013 for GEFCom2014 and 29 March 2016 for Nord Pool and the U.K.) are made.

For models requiring calibration of the regularization parameter (i.e., λ), we use a setup commonly considered in the machine learning literature. Namely, we divide our datasets into estimation (365 days), validation (91 days or 13 full weeks) and test periods (623 days for GEFCom2014, 728 days for Nord Pool and the U.K.; respectively 89 and 104 full weeks). For each of the five models—ridge regression, lasso and elastic nets with $\alpha = 0.25, 0.50$ and 0.75 —34 different ‘sub-models’ with 34 values of λ spanning the regularization parameter space (see Sections 3.4.3 and 3.4.4 for details) are estimated in the 91-day validation period directly following the last day of the initial calibration period; see Figures 1–3. For all hours of the day, only one value of λ is chosen for each of the five models: the one that yields the smallest WMAE error during this 91-day period; for error definitions, see Section 4.1. This value of λ is later used for computing day-ahead price forecasts in the whole out-of-sample test period. To ensure that all models are evaluated using the same data, predictions of all models are compared only in the out-of-sample test periods: 1 April 2012–14 December 2013 (623 days) for GEFCom2014 and 2 April 2014–29 March 2016 (728 days) for Nord Pool and the U.K. Obviously, such a simple procedure for the selection of the regularization parameter may not be optimal. Generally, better performance is to be expected from shrinkage models when λ is recalibrated at every time step. Such an approach has been recently taken by Ziel [18], who used the Bayesian information criterion to select one out of 50 values of λ for every day and every hour in the 969-day-long out-of-sample test period. The downside of such an approach is, however, the increased computational time.

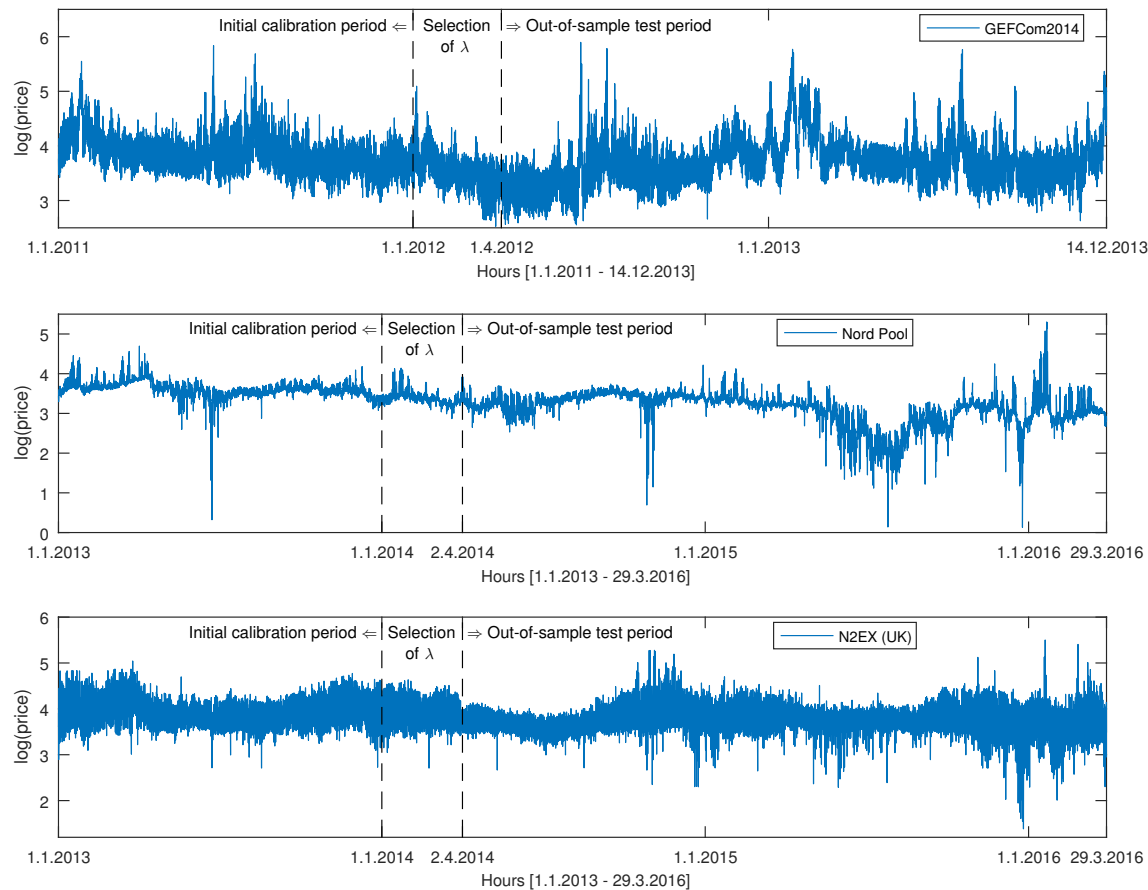


Figure 4. Global Energy Forecasting Competition 2014 (GEFCom2014) (top), Nord Pool (middle) and N2EX (U.K.; bottom) hourly log prices. As expected, the logarithmic transform makes the price series more symmetric. The vertical dashed lines mark the beginning of the 91-day period for selecting λ 's (ridge regression, lasso, elastic nets) and the beginning of the out-of-sample test periods. Each day, the 365-day-long calibration window is rolled forward by 24 h; the models are re-estimated; and price forecasts for the 24 h of the next day are computed.

Our choice of the model classes is guided by the existing literature on short-term EPF. Like in [12,18,25–27,30], the modeling is implemented separately across the hours, leading to 24 sets of parameters for each day the forecasting exercise is performed. As Ziel [18] notes, when we compare the forecasting performance of relatively simple models implemented separately across the hours and jointly for all hours (like in [9,34–36]), the latter generally performs better for the first half of the day, whereas the former are better in the second half of the day. At the same time, models implemented separately across the hours offer more flexibility by allowing for time-varying cross-hour dependency in a straightforward manner. Hence, our choice of the modeling framework.

In the remainder of this section, we first define the benchmarks: a simple similar-day technique and a collection of parsimonious autoregressive models. Since the latter are usually built on some prior knowledge of experts, like in [18], we refer to them as expert models. Then, we move on to describe the selection and shrinkage procedures used in this study.

3.1. The Naive Benchmark

The first benchmark, most likely introduced to the EPF literature in [34] and dubbed the naive method, belongs to the class of similar-day techniques (for a taxonomy of EPF approaches, see, e.g., [1]). It proceeds as follows: the electricity price forecast for hour h on Monday is set equal to the price for the same hour on Monday of the previous week, and the same rule applies for Saturdays

and Sundays; the electricity price forecast for hour h on Tuesday is set equal to the price for the same hour on Monday, and the same rule applies for Wednesdays, Thursdays and Fridays. As was argued in [34,35], forecasting procedures that are not calibrated carefully fail to outperform the naive method surprisingly often. We denote this benchmark by **Naive**.

3.2. Autoregressive Expert Benchmarks

The second benchmark is a parsimonious autoregressive structure originally proposed by Misiolek et al. [19] and later used in a number of EPF studies [18,20,21,23–27]. Within this model, the centered log-price on day d and hour h , i.e., $p_{d,h}$, is given by the following formula:

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

where the lagged log-prices $p_{d-1,h}$, $p_{d-2,h}$ and $p_{d-7,h}$ account for the autoregressive effects of the previous days (the same hour yesterday, two days ago and one week ago), while $p_{d-1}^{\min} \equiv \min_{h=1,\dots,24}\{p_{d-1,h}\}$ is the minimum of the previous day's 24 hourly log-prices. The exogenous variable $z_{d,h}$ refers to the logarithm of hourly system load or Nordic consumption for day d and hour h (actually, to forecasts made a day before, see Section 2). The three dummy variables— D_{Sat} , D_{Sun} and D_{Mon} —account for the weekly seasonality. Finally, the $\varepsilon_{d,h}$'s are assumed to be independent and identically distributed (i.i.d.) normal variables. We denote this autoregressive benchmark by **ARX1** to reflect the fact that the load (or consumption) forecast is used as the exogenous variable in Equation (2). The corresponding model with $\beta_{h,5} \equiv 0$, i.e., with no exogenous variable, is denoted by **AR1**. The **ARX1** and **AR1** models, as well as all autoregressive structures considered in Sections 3.2 and 3.3, are estimated in this study with least squares (LS), using MATLAB's regress.m function.

In what follows, we also consider two variants of Equation (2) that treat holidays as special days:

$$p_{d,h} = \beta_{h,1}p_{d-1,h} + \beta_{h,2}p_{d-2,h} + \beta_{h,3}p_{d-7,h} + \beta_{h,4}p_{d-1}^{\min} + \beta_{h,5}z_{d,h} + \beta_{h,6}D_{Sat} + \beta_{h,7}D_{Sun} + \beta_{h,8}D_{Mon} + \beta_{h,9}D_{Hol} + \varepsilon_{d,h}, \quad (3)$$

and that additionally utilize the fact that prices for early morning hours depend more on the previous day's price at midnight, i.e., $p_{d-1,24}$, than on the price for the same hour, as recently noted in [18,29]:

$$p_{d,h} = \beta_{h,1}p_{d-1,h} + \beta_{h,2}p_{d-2,h} + \beta_{h,3}p_{d-7,h} + \beta_{h,4}p_{d-1}^{\min} + \beta_{h,5}z_{d,h} + \beta_{h,6}D_{Sat} + \beta_{h,7}D_{Sun} + \beta_{h,8}D_{Mon} + \beta_{h,9}D_{Hol} + \beta_{h,10}p_{d-1,24} + \varepsilon_{d,h}. \quad (4)$$

We denote Models (3) and (4) by **ARX1h** and **ARX1hm**, respectively. Similarly, corresponding models with $\beta_{h,5} \equiv 0$ are denoted by **AR1h** and **AR1hm**. Note, that when forecasting the electricity price for the last load period of the day, i.e., $p_{d,24}$, models with suffix **hm** reduce to models with suffix **h** (this is true for all models considered in Section 3.2).

In Equations (3) and (4), D_{Hol} is a dummy variable for holidays. The holidays were identified using the Time and Date AS (www.timeanddate.com/holidays) web page: U.S. federal holidays (for GEFCom2014), national holidays in Norway (for Nord Pool) and public holidays, bank holidays and major observances in the U.K. (option 'Holidays and some observances').

The third benchmark is an extension of the **ARX1** model, which takes into account the experience gained during the GEFCom2014 competition that it may be beneficial to use different model structures for different days of the week, not only different parameter sets [29]. Hence, the multi-day ARX model (denoted later in the text by **mARX1**) is given by the following formula:

$$p_{d,h} = \left(\sum_{i \in I} \beta_{h,1,i} D_i \right) p_{d-1,h} + \beta_{h,2}p_{d-2,h} + \beta_{h,3}p_{d-7,h} + \beta_{h,4}p_{d-1}^{\min} + \beta_{h,5}z_{d,h} + \beta_{h,6}D_{Sat} + \beta_{h,7}D_{Sun} + \beta_{h,8}D_{Mon} + \beta_{h,11}D_{Mon}p_{d-3,h} + \varepsilon_{d,h}, \quad (5)$$

where $I \equiv \{0, \text{Sat}, \text{Sun}, \text{Mon}\}$, $D_0 \equiv 1$ and the term $D_{\text{Mon}}p_{d-3,h}$ accounts for the autoregressive effect of Friday's prices on the prices for the same hour on Monday. Note that to some extent, this structure resembles periodic autoregressive moving average (PARMA) models, which have seen limited use in EPF [37,38]. Like for the **ARX1** model, also for **mARX1**, we consider two variants:

- **mARX1h**, which treats holidays as special days, i.e., with the $\beta_{h,9}D_{\text{Hol}}$ term in Equation (5),
- and **mARX1hm**, which additionally implements the dependence on the previous day's price at midnight, i.e., with the $\beta_{h,9}D_{\text{Hol}}$ and $\beta_{h,10}p_{d-1,24}$ terms in Equation (5).

The corresponding price only models, i.e., with $\beta_{h,5} \equiv 0$, are denoted by **mAR1**, **mAR1h** and **mAR1hm**.

Misiorek et al. [19] noted that the minimum of the previous day's 24 hourly prices was the best link between today's prices and those from the entire previous day. Their analysis, however, was limited to one small dataset (California CalPXprices, 3–9 April 2000) and only one simple function at a time (maximum, minimum, mean or median of the previous day's prices). To check if using more than one function leads to a better forecasting performance, we introduce a benchmark, which is an extension of the **ARX1** model that takes into account not only the minimum (p_{d-1}^{\min}), but also the maximum (p_{d-1}^{\max}) and the mean (p_{d-1}^{avg}) of the previous day's 24 hourly prices. Additionally, we include a second exogenous variable ($y_{d,h}$), which is taken as either the logarithm of the day-ahead zonal load forecast (GEFCom2014) or of the Danish wind power prognosis. The resulting **ARX2** model is given by the following formula:

$$\begin{aligned} p_{d,h} = & \beta_{h,1}p_{d-1,h} + \beta_{h,2}p_{d-2,h} + \beta_{h,3}p_{d-7,h} + \beta_{h,4}p_{d-1}^{\min} + \beta_{h,5}z_{d,h} \\ & + \beta_{h,6}D_{\text{Sat}} + \beta_{h,7}D_{\text{Sun}} + \beta_{h,8}D_{\text{Mon}} \\ & + \beta_{h,11}p_{d-1}^{\max} + \beta_{h,12}p_{d-1}^{\text{avg}} + \beta_{h,13}y_{d,h} + \varepsilon_{d,h}. \end{aligned} \quad (6)$$

Like for the **ARX1** and **mARX1** models, also for **ARX2**, we consider two variants:

- **ARX2h** with the $\beta_{h,9}D_{\text{Hol}}$ term in Equation (6),
- and **ARX2hm** with the $\beta_{h,9}D_{\text{Hol}}$ and $\beta_{h,10}p_{d-1,24}$ terms in Equation (6).

The corresponding price only models, i.e., with $\beta_{h,5}, \beta_{h,13} \equiv 0$, are denoted by **AR2**, **AR2h** and **AR2hm**.

3.3. Full Autoregressive Model

Finally, we define a much richer autoregressive model that includes as special cases all expert models discussed in Section 3.2 and call it the **full ARX** or **fARX** model. We consider all regressors that, in our opinion, possess a non-negligible predictive power. The **fARX** model is similar in spirit to the general autoregressive model defined by Equation (2) in [18]. However, there are some important differences between them. On one hand, **fARX** includes exogenous variables and a much richer seasonal structure. On the other, it does not look that far into the past and concentrates only on days $d-1$, $d-2$, $d-3$ and $d-7$. The **fARX** model is given by the following formula:

$$\begin{aligned} p_{d,h} = & \sum_{i=1}^{24} (\beta_{h,i}p_{d-1,i} + \beta_{h,i+24}p_{d-2,i} + \beta_{h,i+48}p_{d-3,i}) + \beta_{h,73}p_{d-7,h} \\ & + \sum_{j=1}^3 (\beta_{h,j+73}p_{d-j}^{\min} + \beta_{h,j+76}p_{d-j}^{\max} + \beta_{h,j+79}p_{d-j}^{\text{avg}}) \\ & + \beta_{h,83}z_{d,h} + \beta_{h,84}z_{d-1,h} + \beta_{h,85}z_{d-7,h} + \beta_{h,86}y_{d,h} \\ & + \sum_{k=1}^7 \beta_{h,86+k}D_k + \sum_{k=1}^7 \beta_{h,93+k}D_k z_{d,h} + \sum_{k=1}^7 \beta_{h,100+k}D_k p_{d-1,h} + \varepsilon_{d,h}, \end{aligned} \quad (7)$$

where $D_1 \equiv D_{\text{Sat}}, D_2 \equiv D_{\text{Sun}}, \dots, D_7 \equiv D_{\text{Fri}}$ are dummies for the seven days of the week (we treat holidays as the eighth day of the week, hence $D_1 = \dots = D_7 = 0$ for holidays). The price only variant,

fAR, is obtained by setting to zero all coefficients of the terms involving exogenous variables, i.e., $\beta_{h,i} \equiv 0$, for $i = 83, \dots, 86, 94, \dots, 100$.

Although we fit the **fARX** model to power market data and evaluate its forecasting performance, the main reason for including it in this study is to use it as the baseline model for the selection and shrinkage procedures discussed in Section 3.4. For this purpose, let us write the **fARX** model in a more compact form:

$$p_{d,h} = \sum_{i=1}^n \beta_{h,i} X_{d,h,i} + \varepsilon_{d,h}, \quad (8)$$

where $X_{d,h,i}$'s are the $n = 107$ regressors in Equation (7) and $\beta_{h,i}$'s are their coefficients.

3.4. Selection and Shrinkage Procedures

All autoregressive models considered in Sections 3.2 and 3.3 are estimated in this study with least squares (LS). However, there are many alternatives to using LS in multi-parameter models, in particular [39]:

- variable or subset selection, which involves identifying a subset of predictors that we believe to be influential, then fitting a model using LS on the reduced set of variables,
- shrinkage (also known as regularization), which fits the full model with all predictors using an algorithm that shrinks the estimated coefficients towards zero, which can significantly reduce their variance.

Depending on what type of shrinkage is performed, some of the coefficients may be shrunk to zero itself. As such, some shrinkage methods, like the lasso, de facto perform variable selection. It should be noted, however, that variable selection (or model sparsity) is beneficial for interpretability and faster simulation of model trajectories; for reducing the forecasting errors, only the shrinkage property is required.

3.4.1. Single-Step Elimination of Insignificant Predictors

This subset selection procedure is a simple alternative to stepwise regression discussed in Section 3.4.2 and has been used, for instance, in [5]. The idea is to fit the full regression model, in our case **fARX**, then in a single step, set to zero all statistically insignificant coefficients. We use MATLAB's regress.m function with the commonly-used 5% significance level. Setting to zero all coefficients in Equation (7) whose 95% confidence intervals (CI) include zero yields the **ssARX** model for a particular day and hour (the **ssAR** model is obtained analogously from **fAR**; see Section 3.3). This procedure can be conducted by imposing some additional constraints, for instance, leaving in the model all coefficients of the basic **ARX1** (or **AR1**) benchmark. This yields the **ssARX1** and **ssAR1** models. Of course, the most commonly-used significance level of 5% may not be optimal. We have additionally checked the performance of 90% and 97.5% CI. It turns out that the overall ranking of the **ssAR**-type models does not change much. However, **ssARX** and **ssAR** perform slightly better for the 90% CI, while **ssARX1** and **ssAR1** either for the 95% or the 97.5% CI.

3.4.2. Stepwise Regression

Although very fast, the single-step elimination may remove too many explanatory variables at once and lead to a poorly-performing subset of predictors. On the other hand, selecting the best subset from among all 2^n subsets of the n predictors is not computationally feasible for large n . Even if doable, it may lead to overfitting. For these reasons, stepwise methods, which explore a far more restricted set of models, are attractive alternatives to best subset selection [39]. In the context of EPF, they have been used, for instance, in [12,13,40].

There are two basic procedures: forward selection and backward elimination. Forward stepwise selection begins with a model containing no predictors and then iteratively adds variables to the model. At each step, the variable that gives the greatest additional improvement to the fit is added to the

model, and the procedure continues until all important predictors are in the model. We use MATLAB's `stepwisefit.m` function, which computes the p -value of an F -statistic at each time step to test models with and without a potential term. If a variable is not currently in the model, the null hypothesis is that it would have a zero coefficient if added to the model. If there is sufficient evidence to reject the null hypothesis, that variable may be added to the model (we use `stepwisefit`'s default 5% significance level for adding variables; naturally, this could be further fine tuned as for the single-step elimination procedures). In a given step, the function adds the variable with the smallest p -value. We denote the resulting models by **fsARX** and **fsAR**.

Backward stepwise elimination (or selection) begins with the full model containing all n variables, i.e., **fARX** or **fAR**, and then iteratively removes the least useful predictor, one at a time. MATLAB's `stepwisefit.m` function computes the null hypothesis that a given variable has a zero coefficient. If there is insufficient evidence to reject the null hypothesis, the variable may be removed from the model (we use `stepwisefit`'s default 10% significance level for removing variables). In a given step, the function removes the variable with the largest p -value. We denote the resulting models by **bsARX** and **bsAR**.

3.4.3. Ridge Regression

Ridge regression is a regularization method introduced in statistics by Hoerl and Kennard [41]. To our best knowledge, apart from a limited study of Barnes and Balda [16] in the context of evaluating the profitability of battery storage, the method has not been used for EPF. Ridge regression is very similar to least squares, except that the β_i 's in (8) are not estimated by minimizing the residual sum of squares (RSS), but by RSS penalized by a quadratic shrinkage factor:

$$\hat{\beta}^{ridge} = \underset{\beta_{h,i}}{\operatorname{argmin}} \left\{ \text{RSS} + \lambda \sum_{i=1}^n \beta_{h,i}^2 \right\} \equiv \underset{\beta_{h,i}}{\operatorname{argmin}} \left\{ \sum_{d,h \in T} \left(p_{d,h} - \sum_{i=1}^n \beta_{h,i} X_{d,h,i} \right)^2 + \lambda \sum_{i=1}^n \beta_{h,i}^2 \right\}, \quad (9)$$

where T represents the calibration period and $\lambda \geq 0$ is a tuning or regularization parameter, to be determined separately. Note that for $\lambda = 0$, we get the standard LS estimator; for $\lambda \rightarrow \infty$, all $\beta_{h,i}$'s tend to zero; while for intermediate values of λ , we are balancing two ideas: minimizing the RSS and shrinking the coefficients towards zero (and each other).

Ridge regression produces a different set of coefficient estimates for each value of λ ; hence, selecting a good value for λ is critical. Cross-validation provides a simple way to tackle this problem [39]. We choose a grid of λ values (here: 34 equally-spaced values spanning the range from 1–100; if $\lambda \in \{94, 97, 100\}$ was selected, we additionally checked another set of 34 equally-spaced values spanning the range from 101–200) and using MATLAB's `ridge.m` function (we scale the regressors) compute the prediction errors for each value of the tuning parameter in the 91-day validation period; see Section 2. We then select λ for which the WMAE error (for the definition, see Section 4.1) is the smallest and use it for computing day-ahead price forecasts in the whole out-of-sample test period. The resulting model is denoted in the text by **RidgeX** or **Ridge** when the baseline model is **fAR**.

3.4.4. Lasso and Elastic Nets

Ridge regression has one unwanted feature when it comes to interpretation and model identification. Unlike stepwise regression, which will generally select models that involve just a subset of the variables, ridge regression will include all n predictors in the final model [39]. The quadratic shrinkage factor in Equation (9) will shrink all $\beta_{h,i}$'s towards zero, but it will not set any of them exactly to zero. In 1996, Tibshirani [42] proposed the least absolute shrinkage and selection operator (i.e., lasso or LASSO) that overcomes this disadvantage. It is the only shrinkage procedure that has been applied in EPF to a larger extent, however only in the last two years [7,9,18,25,43].

The lasso is a shrinkage method just like ridge regression. However, it uses a linear penalty factor instead of a quadratic one:

$$\hat{\beta}^{lasso} = \underset{\beta_{h,i}}{\operatorname{argmin}} \left\{ \text{RSS} + \lambda \sum_{i=1}^n |\beta_{h,i}| \right\}. \quad (10)$$

This subtle change makes the solutions nonlinear in $p_{d,h}$, and there is no closed form expression as in the case of ridge regression. Because of the nature of the shrinkage factor in Equation (10), making λ sufficiently large will cause some of the coefficients to be exactly zero [44]. Thus, the lasso de facto performs variable selection, just like the methods discussed in Sections 3.4.1 and 3.4.2. As in ridge regression, selecting a good value of λ for the lasso is critical. Here, we use MATLAB's `lasso.m` function and a grid of exponentially-decreasing λ 's (the largest just sufficient to produce all $\beta_i = 0$; the function also automatically scales the regressors). We then select λ for which the $\overline{\text{WMAE}}$ error (for the definition, see Section 4.1) in the 91-day validation period is the smallest. The resulting model is denoted in the text by **LassoX**, or **Lasso** when the baseline model is **fAR**.

The lasso does not handle highly-correlated variables very well. The coefficient paths tend to be erratic and can sometimes show wild behavior [44]. This is not a critical issue for forecasting, but for interpretation and model identification, this has more serious consequences. In 2005, Zou and Hastie [45] proposed the elastic net, a new regularization and variable selection method that can be seen as an extension of ridge regression and the lasso. It often outperforms the lasso, while exhibiting a similar sparsity of representation. The elastic net uses a mixture of linear and quadratic penalty factors:

$$\hat{\beta}^{EN} = \underset{\beta_{h,i}}{\operatorname{argmin}} \left\{ \text{RSS} + \lambda \left(\frac{1-\alpha}{2} \sum_{i=1}^n \beta_{h,i}^2 + \alpha \sum_{i=1}^n |\beta_{h,i}| \right) \right\}, \quad (11)$$

where $\alpha \in [0, 1]$. When $\alpha = 1$, the elastic net reduces to the lasso, and with $\alpha = 0$, it becomes ridge regression. The $\frac{1}{2}$ in the quadratic part of the elastic net penalty in Equation (11) leads to a more efficient and intuitive soft-thresholding operator in the optimization; the original formulation in [45] did not include the $\frac{1}{2}$ scaling. Note also that every elastic net problem can be rewritten as a lasso problem on augmented data. Hence, for fixed λ and α , the computational difficulty of the elastic net solution is similar to the lasso problem [44].

Compared to the lasso and ridge regression, the elastic net has an additional mixing parameter that has to be determined. It can be set on subjective grounds, as we do here, or optimized within a cross-validation scheme. We use MATLAB's `lasso.m` function (with a grid of exponentially-decreasing λ 's; the function also automatically scales the regressors) and three values of the mixing parameter, $\alpha = 0.25, 0.50$ and 0.75 . This yields six elastic net models:

- **EN25X**, **EN50X** and **EN75X** when the baseline model is **fARX**,
- and **EN25**, **EN50** and **EN75** when the baseline model is **fAR**,

that span the space between ridge regression (**RidgeX**, **Ridge**) and lasso models (**LassoX**, **Lasso**).

4. Empirical Results

We now present day-ahead forecasting results for the three considered datasets: GEFCom2014 hourly locational marginal prices, Nord Pool hourly system prices and U.K. hourly system prices. We use long, two-year out-of-sample test periods to make sure the obtained results are reliable (for the GEFCom2014 dataset, the test period is shorter: 623 days; see Figure 1). Recall from Section 2 that the models are re-estimated on a daily basis. Price forecasts $\hat{P}_{d+1,1}, \dots, \hat{P}_{d+1,24}$ for all 24 h of the next day are determined at the same point in time, and the 365-day calibration window is rolled forward by one day.

4.1. Performance Evaluation in Terms of WMAE

Following [21,24,30,35], we compare the models in terms of the weekly-weighted mean absolute error (WMAE) loss function, which is a robust measure similar to MAPE, but with the absolute error normalized by the mean weekly price to avoid the adverse effect of negative and close to zero electricity spot prices. We evaluate the forecasting performance using weekly time intervals, each with $24 \times 7 = 168$ hourly observations. For each week $w = 1, \dots, w_{max}$ in the out-of-sample test period, we calculate the error for each model as:

$$\text{WMAE}_w = \frac{1}{\bar{P}_{168}} \text{MAE}_w = \frac{1}{168 \cdot \bar{P}_{168}} \sum_{d=\text{Mon}}^{\text{Sun}} \sum_{h=1}^{24} |P_{d,h} - \hat{P}_{d,h}|, \quad (12)$$

where $P_{d,h}$ is the actual price for hour h (not the centered log-price $p_{d,h}$), $\hat{P}_{d,h}$ is the model predicted price for that hour, $\bar{P}_{168} = \frac{1}{168} \sum_{d=\text{Mon}}^{\text{Sun}} \sum_{h=1}^{24} P_{d,h}$ is the mean price for a given week and $w_{max} = 89$ for GEFCom2014 and 104 for Nord Pool and the U.K. Next, we aggregate these errors into one mean value over all weeks in the out-of-sample test period:

$$\overline{\text{WMAE}} = \frac{1}{w_{max}} \sum_{w=1}^{w_{max}} \text{WMAE}_w. \quad (13)$$

Note that we also analyzed the forecasts using the weekly root mean square error (see [1] (Section 3.3)), but the results were qualitatively the same and are omitted here due to space limitations.

In Table 1, we report $\overline{\text{WMAE}}$ errors for the three considered datasets and the 20 model types. We use the GEFCom2014 dataset twice: once we fit ARX-type models to the complete dataset with exogenous variables (system and zonal load; left part of the table) and once we fit AR-type models to the dataset without them (right part of the table). This allows us to compute the decrease in WMAE when exogenous variables are added to the model (the last column in Table 1). Several important conclusions can be drawn:

- All models beat the **Naive** benchmark and, except for the **fAR** model and the U.K. data, by a large margin. In particular, the improvement from using elastic nets can be as much as 5%! This indicates that they all are highly efficient forecasting tools.
- When we exclude single-step elimination without constraints (**ssAR/X**) and backward selection (**bsAR/X**) models, the selection and shrinkage methods generally outperform the expert benchmarks. In particular, the elastic net model with $\alpha = 0.75$ (i.e., closer in terms of α to the lasso than to ridge regression) beats every expert model, except **mAR1hm** for the U.K. data, where it is second best.
- The latter comment leads us to the next conclusion that adding the price for the last load period of the day, $p_{d-1,24}$, to the expert models improves their performance greatly. This fact has been recognized in the EPF literature only very recently [18,25,29] and apparently requires more attention. To see this, compare the models with suffix **m** to those without it. In particular, **mAR1hm** is the overall best performing model for the U.K. dataset and **ARX2hm** is the third best model for the Nord Pool dataset.
- Somewhat surprisingly, the full ARX model performs poorly. For the U.K. dataset, it is nearly as bad as the **Naive** benchmark. In all four cases (three datasets + GEFCom2014 without exogenous variables), it is worse than the overall best model and the best performing elastic net (**EN75/X**) by at least 1.4%. Given that a 1% improvement in MAPE translates into savings of ca. \$1.5 million per year for a typical medium-size utility [2,3], this observation is of high practical value. Yet, from a statistical perspective, this finding is not that surprising. The **fARX** model has 107 parameters, which have to be calibrated to only 365 observations. Increasing the length of the calibration window should lead to a better performance of the full model.

Table 1. Mean values of the weekly-weighted mean absolute errors, i.e., $\overline{\text{WMAE}}$ defined by Equation (13), over all 89 weeks of the GEFCom2014 or all 104 weeks of the Nord Pool and U.K. out-of-sample test periods. WMAE errors are reported in percent, with standard deviation in parentheses. A heat map is used to indicate better (\rightarrow green) and worse (\rightarrow red) performing models. $\overline{\text{WMAE}}$ errors for the best performing model for each dataset are emphasized in bold. The last column presents the decrease in $\overline{\text{WMAE}}$ when exogenous variables are added to the model ($\text{AR} \rightarrow \text{ARX}$; for the GEFCom2014 dataset). The bottom rows compare the performance across model classes.

	ARX-type			AR-type		AR - ARX
	GEFCom	Nord Pool		GEFCom	N2EX (UK)	GEFCom
Naive	14.708 (0.975)	11.141 (0.778)	Naive	14.708 (0.975)	9.767 (0.310)	0.000
Expert benchmarks						
ARX1	11.069 (0.639)	9.739 (0.614)	AR1	11.183 (0.701)	8.384 (0.253)	0.114
ARX1h	11.072 (0.639)	9.693 (0.616)	AR1h	11.181 (0.704)	8.389 (0.253)	0.109
ARX1hm	10.976 (0.617)	8.673 (0.516)	AR1hm	11.062 (0.657)	8.229 (0.247)	0.086
mARX1	11.102 (0.621)	9.482 (0.601)	mAR1	11.320 (0.696)	8.258 (0.253)	0.218
mARX1h	11.105 (0.622)	9.461 (0.602)	mAR1h	11.322 (0.699)	8.270 (0.254)	0.218
mARX1hm	10.974 (0.598)	8.461 (0.518)	mAR1hm	11.168 (0.644)	8.098 (0.246)	0.195
ARX2	10.742 (0.575)	8.878 (0.546)	AR2	11.331 (0.700)	8.290 (0.253)	0.589
ARX2h	10.739 (0.575)	8.826 (0.546)	AR2h	11.333 (0.704)	8.288 (0.253)	0.594
ARX2hm	10.625 (0.565)	8.206 (0.485)	AR2hm	11.070 (0.656)	8.237 (0.249)	0.444
Full ARX model						
fARX	10.911 (0.507)	10.131 (0.708)	fAR	12.279 (0.602)	9.724 (0.334)	1.368
Selection and shrinkage methods						
ssARX	10.669 (0.577)	8.861 (0.537)	ssAR	12.061 (0.644)	9.344 (0.270)	1.393
ssARX1	9.894 (0.548)	8.409 (0.507)	ssAR1	11.343 (0.641)	8.395 (0.261)	1.449
fsARX	9.876 (0.502)	8.130 (0.502)	fsAR	11.193 (0.592)	8.563 (0.272)	1.317
bsARX	10.449 (0.502)	9.421 (0.599)	bsAR	11.968 (0.582)	9.252 (0.301)	1.519
RidgeX	9.777 (0.544)	8.972 (0.479)	Ridge	10.775 (0.653)	8.237 (0.260)	0.998
LassoX	9.476 (0.516)	8.419 (0.503)	Lasso	10.722 (0.609)	8.125 (0.253)	1.246
EN75X	9.475 (0.517)	8.056 (0.489)	EN75	10.708 (0.610)	8.124 (0.253)	1.233
EN50X	9.473 (0.518)	8.287 (0.496)	EN50	10.688 (0.611)	8.121 (0.253)	1.215
EN25X	9.474 (0.522)	8.529 (0.503)	EN25	10.650 (0.613)	8.113 (0.253)	1.176
Comparisons						
Expert - Best	1.152	0.150	Expert - Best	0.412	0.000	
fARX - Best	1.438	2.075	fAR - Best	1.629	1.626	
Naive - Best	5.235	3.086	Naive - Best	4.058	1.670	

- Among the selection and shrinkage methods, the lasso and elastic nets tend to outperform single-step elimination (**ssAR/X/1**), stepwise regression (**fsAR/X**, **bsAR/X**) and even ridge regression (**Ridge/X**). Only for the Nord Pool dataset, the **fsARX** forward selection model is better than the lasso and two elastic nets.

4.2. Diebold–Mariano Tests

In order to formally investigate the advantages from using selection and shrinkage methods, we apply the Diebold–Mariano (DM; see [31]) test for significant differences in the forecasting performance. Since predictions for all 24 h of the next day are made at the same time using the same information set, forecast errors for a particular day will typically exhibit high serial correlation. Therefore, like [24,30,46], we conduct the DM tests for each of the 24 load periods separately, using absolute error losses of the model forecast:

$$L(\varepsilon_t) = |\varepsilon_t| = |P_{d,h} - \hat{P}_{d,h}|. \quad (14)$$

For each pair of models and for each hour independently, we calculate the loss differential series:

$$d_t = L(\varepsilon_t^{model_X}) - L(\varepsilon_t^{model_Y}). \quad (15)$$

We perform two one-sided DM tests at the 5% significance level: (i) a test with the null hypothesis $H_0: E(d_t) \leq 0$, i.e., the outperformance of the forecasts of $model_Y$ by those of $model_X$; and (ii) the complementary test with the reverse null $H_0^R: E(d_t) \geq 0$, i.e., the outperformance of the forecasts of $model_X$ by those of $model_Y$. Note that, like in [24,30,46], we assume here forecasts for consecutive days, hence loss differentials are not serially correlated. For the better performing models, this is a generally valid assumption.

In Figures 5 and 6, we summarize the DM results for all test cases (three datasets + GEFCom2014 without exogenous variables). Namely, we sum the number of significant differences in forecasting performance across the 24 h and use a heat map to indicate the number of hours for which the forecasts of a model on the X-axis are significantly better than those of a model on the Y-axis. Two extreme cases—(i) the forecasts of a model on the X-axis are significantly better for all 24 h of the day and (ii) the forecasts of a model on the X-axis are not significantly better for any hour—are indicated by white and black squares, respectively. Naturally, the diagonal (white crosses on black squares) should be ignored, as it concerns the same model on both axes. Columns with many non-black squares (the more green or white the better) indicate that the forecasts of a model on the X-axis are significantly better than the forecasts of many of its competitors. Conversely, rows with many non-black squares mean that the forecasts of a model on the Y-axis are significantly worse than the forecasts of many of its competitors. For instance, for the GEFCom2014 dataset and ARX-type models displayed in the left panel of Figure 5, the white row for the **Naive** benchmark indicates that the forecasts of this simple model are significantly worse than the forecasts of all of its competitors for all 24 h, while the black column for the **Naive** benchmark means that not a single competitor produces significantly worse forecasts than **Naive**, even for a single hour of the day.

The obtained DM-test results support our observations from Section 4.1 on WMAE errors. Again, we can conclude that applying the lasso or one of the elastic nets improves forecasting accuracy. Especially for the GEFCom2014 dataset (both for ARX- and AR-type models), these variable selection schemes lead to models that yield significantly better forecasts than those of the expert models (see the white columns for **Lasso/X**, **EN75/X**, **EN50/X** and **EN25/X** in the left panels of Figures 5 and 6), while their predictions are never outperformed by any of the competitors (see the black rows for these four models). For the Nord Pool and U.K. datasets, the results are not that clear cut, but still there are many more green or white squares in the columns than in the rows corresponding to these four selection schemes.

Again, **EN75/X** stands out as the best performing model overall. For the GEFCom2014 test cases, it always leads to significantly better forecasts than any of the expert benchmarks. For the Nord Pool dataset, its forecasts are significantly better for 10–23 h of the day and significantly worse for at most 2 h (only for models with suffix **m**: **mARX1hm** and **ARX2hm**, 2 h, and **ARX1hm**, 1 h). Finally, for the U.K. dataset, the results are the least convincing. **EN75/X** yields significantly better forecasts for 4–12 h of the day and significantly worse for at most 2 h (only for **mAR**-type models: **mAR1** and **mAR1h**, 2 h, and **mAR1hm**, 1 h).

Now, let us look in detail at the performance for each hour of the day. In Figure 7, we provide a graphical representation of the DM test statistic for four models and all considered datasets. The models include: the best overall **EN75/X** model and three benchmarks (**Naive**, **mARX1hm/mAR1hm** and **fAR/X**). For the GEFCom2014 dataset, **EN75/X** clearly beats all benchmarks across all hours. The situation for the remaining two datasets would be nearly the same if it was not for the early morning hours (Hours 6 and 7 for Nord Pool and Hour 8 for the U.K.), when the expert benchmarks yield significantly better predictions. This is somewhat surprising, since the morning peak comes a bit later in both markets. Perhaps looking at variables selected by the elastic net algorithm will provide more insight.

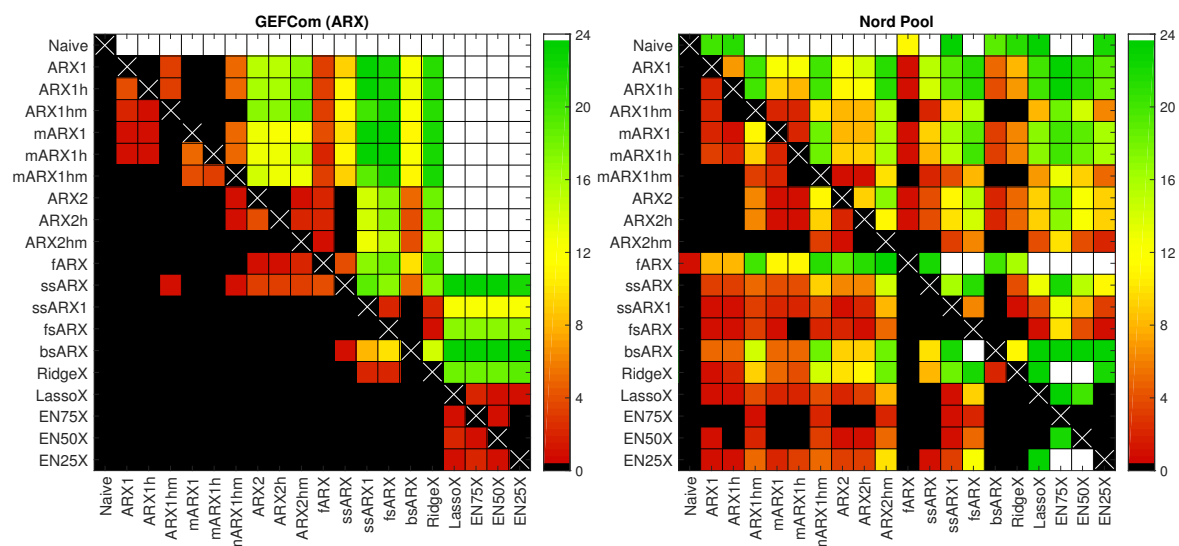


Figure 5. Results for conducted one-sided Diebold–Mariano tests at the 5% level for autoregressive model structures with exogenous variables (ARX)-type models and two datasets: GEFCom2014 (left panel) and Nord Pool (right panel). We sum the number of significant differences in forecasting performance across the 24 h and use a heat map to indicate the number of hours for which the forecasts of a model on the X-axis are significantly better than those of a model on the Y-axis. A white square indicates that forecasts of a model on the X-axis are better for all 24 h, while a black square that they are not better for a single hour.

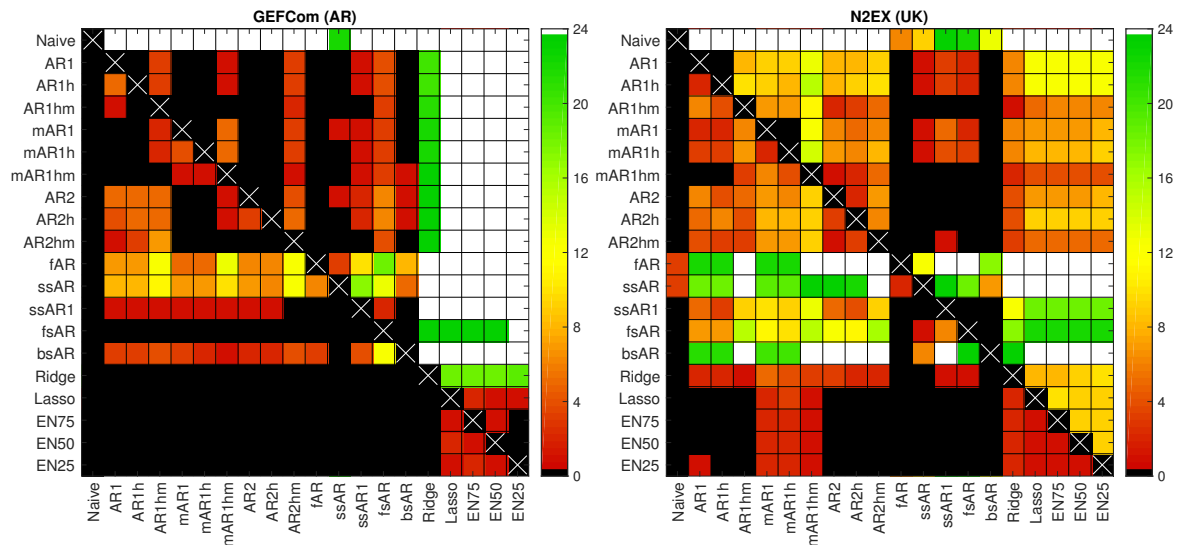


Figure 6. Results for conducted one-sided Diebold–Mariano tests at the 5% level for AR-type models and two datasets: GEFCom2014 (left panel) and N2EX (U.K.; right panel). We sum the number of significant differences in forecasting performance across the 24 h and use a heat map to indicate the number of hours for which the forecasts of a model on the X-axis are significantly better than those of a model on the Y-axis. A white square indicates that forecasts of a model on the X-axis are better for all 24 h, while a black square that they are not better for a single hour.

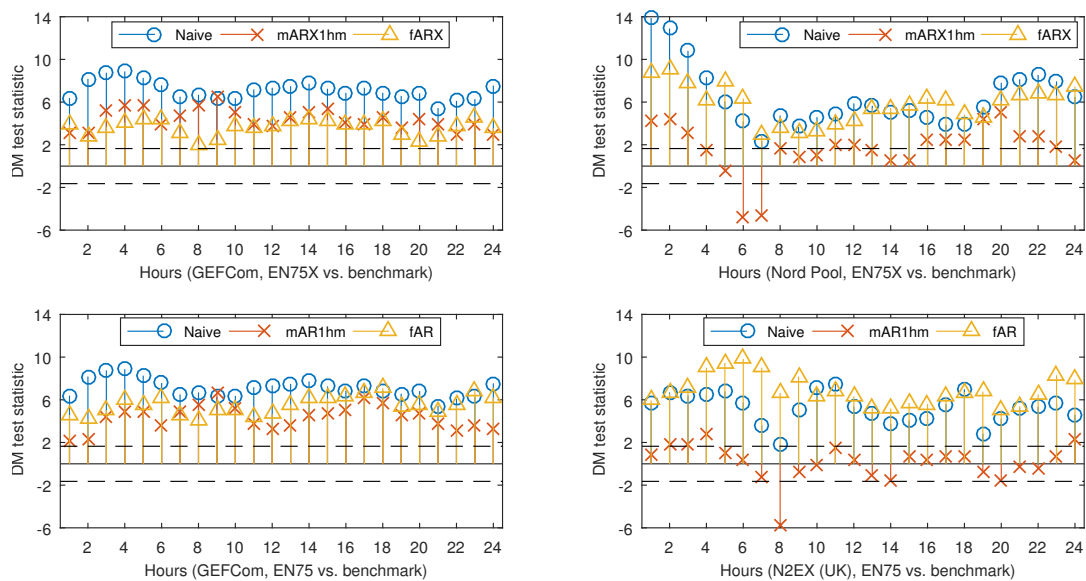


Figure 7. Results for the conducted one-sided Diebold–Mariano tests at the 5% significance level for selected ARX-type models and the GEFCom2014 and Nord Pool datasets (top panels) and selected AR-type models and the GEFCom2014 and N2EX (U.K.) datasets (bottom panels). The tests were conducted separately for each of the 24 h. The figures report the value of the test statistic for each test, as well as two thresholds (dashed lines in the plots). The lower one refers to null hypothesis $H_0: E(d_t) \leq 0$, i.e., the outperformance of the forecasts of EN75/X by those of a given benchmark (Naive, mARX1hm/mAR1hm, fAR/X). The upper threshold refers to the complementary test with the reverse null, i.e., $H_0: E(d_t) \geq 0$ or the outperformance of the forecasts of a given benchmark by those of EN75/X. Only points lying below (or above) the dashed threshold lines are significant at the 5% level. See also Figures 5 and 6.

4.3. Variable Selection

In Tables 2 and 3, we provide the number of days in the out-of-sample test period for which a given $\beta_{h,i}$ was selected for the best performing elastic net model, i.e., EN75/X. The maximum number of days is 623 ($= 7 \times 89$ weeks) for GEFCom2014 and 728 ($= 7 \times 104$ weeks) for Nord Pool and N2EX (U.K.). A heat map is used to indicate more (\rightarrow green) and less (\rightarrow red) commonly-selected $\beta_{h,i}$'s. The $\beta_{h,i}$'s are numbered as in Equation (7). Note that $\beta_{h,83}, \dots, \beta_{h,86}, \beta_{h,94}, \dots, \beta_{h,100} \equiv 0$ in the EN75 model; see Table 3. Several interesting conclusions can be drawn:

- There is no single variable that is always used, regardless of the dataset, hour of the day or the day in the out-of-sample test period. The closest to 'perfection' is the day-ahead load forecast for the predicted hour, i.e., $z_{d,h}$ (see Row 83 in Table 2). Surprisingly, this dependence on the load forecast is stronger than the autoregressive effect (see the next bullet point). This may be a hint that the load-price relationship should be given more attention and that functionals of load-related (or other fundamental) variables should be included in EPF models, like in [10].
- As expected, the price 24 h ago, i.e., $p_{d-1,h}$, is an influential variable; see the diagonals in Rows 1–24 in both Tables. However, it is not only the same hour a day earlier, but also the neighboring hours. The diagonal is less visible around mid-day, and for Nord Pool, it almost disappears except for the late night hours. The latter may be to some extent due to the importance of wind in this market and the explanatory power of the day-ahead wind prognosis for the predicted hour.
- As recently observed in [18,29], the price for Hour 24, i.e., $p_{d-1,24}$, is an influential variable. Somewhat surprisingly, sometime between 7–9 a.m. and 9–11 p.m., Hour 22, i.e., $p_{d-1,22}$, becomes more important. What is more surprising, these late night hours are generally more often selected than the same hour a day ago, i.e., $p_{d-1,h}$. These observations require more thorough studies. Nevertheless, our limited results suggest that these late hour variables should be taken into account when constructing expert models.
- Clearly, the least important variables for all markets are the daily average prices over the last three days, i.e., p_{d-j}^{avg} for $j = 1, 2, 3$, which are almost never selected. There are some exceptions, though, for the GEFCom2014 dataset and the EN75 model; see Table 3. Of the two other aggregated variables, p_{d-j}^{max} is slightly more influential than p_{d-j}^{min} , which contradicts the observations of Misiorek et al. [19] and may suggest its use in expert models instead of the minimum.
- If prices from days $(d-2)$ or $(d-3)$ are ever selected, it is only for hours around midnight the day before (i.e., $p_{d-2,23}$, $p_{d-2,24}$, $p_{d-3,1}$) or similar hours (i.e., the diagonals in Rows 25–48 and 49–72). On the other hand, the same hour one week ago, i.e., $p_{d-7,h}$, has a high explanatory power (see Row 73 for all datasets), which justifies its use in expert models [18–23,30].
- Finally, the weekly dummies (Rows 87–93), the dummy-linked load forecasts (Rows 94–100 in Table 2 only) and the dummy-linked last day's prices (Rows 101–107) are generally selected for the EN75/X model. This may be an indication that the weekly seasonality requires better modeling than offered by typically-used expert models.

Table 2. The number of days in the out-of-sample test period for which a given $\beta_{h,i}$ (see Equation (7)) was selected for the EN75X model. The maximum number of days is 623 ($= 7 \times 89$ weeks) for GECom2014 and 728 ($= 7 \times 104$ weeks) for Nord Pool. A heat map is used to indicate more (\rightarrow green) and less (\rightarrow red) commonly-selected $\beta_{h,i}$'s.

GECom2014																								Nord Pool																										
$\beta_{h,i}$	Hour																								$\beta_{h,i}$	Hour																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1	600	621	314	116	126	522	493	135	316	402	498	490	452	489	456	469	297	411	343	312	452	562	455		1	492	393	284	355	146	295	420	385	358	346	351	354	304	451	392	321	388	397	362	407	248	281	294	24	
2	610	623	471	390	191	352	44	21	11	2	102	78	26	52	13	0	0	44	30	102	112	133	53	51	2	153	115	79	139	100	114	15	10	45	41	39	30	34	28	60	53	49	19	12	20	81	35	129		
3	38	453	580	465	271	164	206	75	106	69	155	158	14	191	154	267	298	171	76	164	21	0	0	0	0	10	363	384	348	313	366	4	48	6	56	35	39	64	53	27	103	125	71	141	238	285	223	497		
4	205	197	390	550	369	184	2	50	10	0	0	0	0	0	0	0	0	9	18	0	0	1	0	0	7	5	19	57	73	40	39	0	17	73	110	80	99	22	7	8	6	6	5	0	24	21	54	16	36	
5	289	239	396	460	621	282	101	180	41	201	293	154	81	17	1	38	0	62	210	0	128	26	6	0	0	5	315	260	481	416	409	387	305	230	206	140	74	43	132	127	148	112	95	63	138	189	165	340		
6	307	467	596	618	604	623	566	370	184	115	284	316	336	337	257	336	138	147	176	51	233	452		6	323	364	359	484	495	365	421	311	194	250	343	396	484	341	382	262	376	335	210	299	309	442				
7	8	80	205	158	329	197	623	623	595	613	399	264	80	16	40	15	51	216	317	507	286	326	313	130	8	85	442	438	383	445	550	282	273	270	272	256	350	301	239	332	380	366	243	116	131	190	32	162		
8	119	85	164	205	237	275	406	619	535	422	151	41	58	38	3	90	204	181	103	251	88	19	5	8	11	0	0	0	24	17	298	2	300	250	242	95	32	61	35	17	9	56	57	381	17	239	245			
9	87	0	0	0	6	91	43	67	318	314	73	61	56	28	131	30	0	0	0	43	0	14	45	9	19	129	176	165	135	207	437	251	113	298	179	82	102	124	175	247	173	100	308	381	423	381	295	252		
10	269	3	0	0	2	9	261	112	100	70	58	15	64	170	251	304	315	296	286	280	266	278	277	11	107	45	56	81	56	109	160	105	141	121	61	144	226	220	259	243	339	332	194	158	154	173	7			
11	17	16	17	0	0	0	73	64	242	426	501	556	245	139	7	14	7	123	150	87	1	3	11	0	11	14	57	44	44	191	41	2	357	428	344	355	294	83	141	0	31	146	183	176	78	5	7	0		
12	12	17	10	0	0	0	0	0	0	177	284	263	168	187	162	74	18	0	15	1	0	2	23	12	28	244	244	314	325	95	282	81	50	6	55	121	259	269	195	11	42	4	73	59	28	0	0	0		
13	61	88	0	78	9	0	0	13	0	117	269	305	324	304	158	93	45	54	59	101	83	53	62	36	163	138	155	144	12	10	25	20	0	33	21	76	68	92	102	99	21	28	82	105	15	0	0			
14	13	64	0	0	0	14	35	106	5	0	18	51	206	239	182	109	90	5	78	160	183	122	134	19	143	33	2	15	32	8	57	20	56	30	26	113	110	121	65	31	50	9	0	0	0	13	17			
15	16	0	26	31	244	132	139	198	70	27	40	75	128	280	346	344	356	338	339	121	152	144	20	0	157	185	177	101	126	108	143	102	266	303	307	61	20	115	10	121	65	31	50	9	0	0	13	17		
16	347	372	335	443	340	328	234	216	88	26	0	2	119	112	144	206	154	599	188	136	92	64	106	178	347	408	356	354	266	108	105	202	62	193	155	111	14	225	283	180	175	150	152	63	64	106	178			
17	32	54	20	36	76	468	580	528	566	447	298	212	202	145	257	213	360	121	313	344	361	138	146	179	189	319	502	402	287	93	306	307	140	459	413	424	436	414	395	661	142	148	659	183	0	216	440			
18	352	367	268	253	143	129	1	132	71	130	104	30	48	63	322	439	488	623	329	16	66	37	211	186	124	176	211	369	390	483	442	471	404	489	429	415	436	414	422	411	455	619	481	275	199	238	307			
19	354	475	565	477	421	325	385	576	325	365	360	470	491	575	547	541	323	495	623	369	173	588	416	283	194	277	295	408	488	212	119	38	110	165	148	54	51	48	103	122	48	104	109	121	50	91	60	36		
20	118	239	169	85	72	387	623	452	298	298	362	388	511	455	538	457	407	258	303	623	332	299	362	272	194	275	310	429	456	44	0	180	231	257	226	193	232	196	202	216	252	231	275	260	35	27	31	8		
21	615	561	326	189	247	381	283	623	623	608	623	623	623	623	623	623	623	618	479	556	623	623	623	220	188	126	163	166	179	639	319	112	65	55	51	68	64	82	70	72	53	17	48	391	459	217	3	7		
22	326	324	426	385	387	329	353	600	623	623	623	623	623	623	623	623	623	623	614	619	623	623	623	220	443	425	625	695	696	656	175	651	728	728	728	728	728	721	728	707	657	682	677	683	692	674	375			
23	177	93	227	246	279	238	510	555	574	445	436	340	241	242	265	414	304	359	385	320	372	400	623	571	23	15	86	218	359	279	320	721	553	217	128	341	533	553	596	432	360	405	478	384	292	247	656	600		
24	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	623	220	128	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728	728
25	15	217	287	386	445	201	7	14	16	22	55	54	70	63	103	112	59	47	19	66	101	29	12	66	25	359	374	686	686	551	498	454	254	195	188	189	84	166	363	395	317	303	262	151	187	415	446	496	635	
26	0	20	8	31	35	92	50	81	131	144	239	111	115	100	51	50	47	57	80	3	53	34	12	66	24	124	222	222	686	686	551	498	454	254	195	188	189	84	166	363	395	317	303	262	151	187	415	446	496	635
27	0	0	18	1	44	35	6	125	133	166	76	55	218	224	271	297	337	111	202	83	21	36	24	21	27	0	15	74	11	34	124	24	81	143	114	181	93	15	17	54	128	39	8	90	104	117	155	132	251	
28	8	13	0	0	17	0	0	28	0	49	86	75	95	48	14	12	2	74	141	92	52	53	0	0	10	17	45	48	129	140	156	149	178	78	20	0	1	0	5	1	3	41	47	0	3	0	0			
29	141	233	214	153	176	52	16	0	0	0	0	0	70																																					

Table 3. The number of days in the out-of-sample test period for which a given $\beta_{h,i}$ (see Equation (7)) was selected for the EN75 model. The maximum number of days is 623 ($=7 \times 89$ weeks) for GEFCom2014 and 728 ($=7 \times 104$ weeks) for N2EX (U.K.). A heat map is used to indicate more (\rightarrow green) and less (\rightarrow red) commonly-selected $\beta_{h,i}$'s. Note that $\beta_{h,83}, \dots, \beta_{h,86}, \beta_{h,94}, \dots, \beta_{h,100} \equiv 0$ in the EN75 model.

$\beta_{h,i}$	GEFCom2014																								$\beta_{h,i}$	N2EX(UK)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
	Hour																									Hour																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
1	454	393	0	0	0	81	371	156	126	260	321	389	394	394	376	367	334	250	317	255	257	345	422	448	1	228	291	287	382	374	370	393	115	366	175	137	11	71	84	143	267	67	181	0	142	23	7	0	272																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
2	573	623	383	180	39	169	47	0	0	7	129	120	21	55	44	40	24	58	15	10	1	150	36	117	2	144	724	627	0	67	27	35	105	167	453	194	172	358	439	514	509	156	1	372	353	295	198	104	135																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
3	31	437	493	245	36	20	191	0	7	5	160	115	50	144	192	218	232	219	72	177	103	0	0	0	3	0	282	382	356	116	94	121	249	113	53	113	103	35	30	48	0	0	131	100	4	7	121	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
4	332	559	613	623	396	176	0	184	35	0	20	0	0	19	9	20	30	2	1	0	0	0	0	8	68	4	0	106	226	616	566	418	96	161	140	123	77	34	214	64	51	51	68	0	85	248	92	805	148																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
5	303	285	362	406	617	279	161	207	42	129	184	10	0	0	3	7	5	45	67	6	74	0	0	0	5	66	53	85	183	182	96	161	145	0	51	85	8	1	0	0	323	4	11	214	0	40	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
6	249	492	575	599	623	623	546	334	358	209	145	291	333	374	375	371	282	51	171	93	63	0	1	6	215	117	170	185	298	500	256	89	2	91	44	1	0	118	16	51	2	40	186	194	142	45	91	36																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
7	30	54	51	41	169	17	623	589	537	551	57	86	279	468	618	534	458	322	214	131	265	239	246	258	7	165	57	149	456	226	474	273	285	554	591	443	296	242	45	161	0	0	136	380	60	68	81	0	63																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
8	0	0	0	0	0	59	141	281	576	505	287	0	51	215	178	170	160	107	25	0	22	30	31	147	0	8	430	656	558	520	521	628	509	391	240	337	61	58	82	181	111	0	106	101	0	0	84	179																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
9	8	0	0	0	0	0	0	0	0	27	73	68	38	236	264	298	357	383	308	242	237	25	208	73	0	9	93	389	425	456	433	625	601	547	446	319	75	0	352	200	0	39	66	58	15	0	26	224																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
10	33	0	0	0	0	0	52	0	34	81	46	184	316	416	496	561	559	426	473	493	457	455	472	182	0	10	399	68	0	0	0	0	51	178	728	481	76	333	272	0	0	65	148	267	178	115	339																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
11	0	0	0	0	0	0	0	0	64	57	120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	2	202	326	323	24	120	8	0	117	181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5. Conclusions

A key point in electricity price forecasting (EPF) is the appropriate choice of explanatory variables. The typical approach has been to select predictors in an ad hoc fashion, sometimes using expert knowledge, but very rarely based on formal selection or shrinkage procedures. However, is this the right approach? Can the application of automated selection and shrinkage procedures to large sets of explanatory variables lead to better forecasts than those of the commonly-used expert models?

Conducting an empirical study involving state-of-the-art parsimonious autoregressive structures as benchmarks, datasets from three major power markets and five classes of automated selection and shrinkage procedures (single-step elimination, stepwise regression, ridge regression, lasso and elastic nets), we have addressed these important questions. To this end, we have compared the predictive performance of 20 types of models over three two-year-long out-of-sample test periods in terms of the robust weekly-weighted mean absolute error (WMAE) and tested the statistical significance of the results using the Diebold–Mariano [31] test.

We have shown that two classes of selection and shrinkage procedures—the lasso and elastic nets—lead to on average better performance than any of the considered expert benchmarks. On the other hand, single-step elimination, stepwise regression and ridge regression are not recommended for EPF as they do not yield significant accuracy gains compared to well-structured parsimonious autoregressive models. The lasso has been recently shown to perform well in EPF [9,18], but it is the more flexible elastic net that stands out as the best performing model overall. Given that both are automated procedures that do not require advanced expert knowledge or supervision, our results may have far reaching consequences for the practice of electricity price forecasting.

We have also looked at variables selected by the elastic net algorithm to gain insights for constructing efficient parsimonious models. In particular, we have confirmed the high explanatory power of the load forecasts for the target hour, of last day's prices for the same or neighboring hours and of the price for the same hour a week earlier. Somewhat surprisingly, we have found that not only the last available data point (price for Hour 24), but also prices for Hours 21–23 of the previous day should be considered when building expert models.

Acknowledgments: The study was partially supported by the National Science Center (NCN, Poland) through Grants 2015/17/B/HS4/00334 (to Bartosz Uniejewski and Rafał Weron) and 2013/11/N/HS4/03649 (to Jakub Nowotarski).

Author Contributions: All authors conceived of and designed the forecasting study. Bartosz Uniejewski performed the numerical experiments. Jakub Nowotarski and Rafał Weron supervised the experiments. All authors analyzed the results and contributed to writing the paper.

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.
2. Zareipour, H.; Canizares, C.A.; Bhattacharya, K. Economic impact of electricity market price forecasting errors: A demand-side analysis. *IEEE Trans. Power Syst.* **2010**, *25*, 254–262.
3. Hong, T. Crystal Ball Lessons in Predictive Analytics. *EnergyBiz Mag.* **2015**, 35–37.
4. 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.
5. Gianfreda, A.; Grossi, L. Forecasting Italian electricity zonal prices with exogenous variables. *Energy Econ.* **2012**, *34*, 2228–2239.
6. Maciejowska, K. Fundamental and speculative shocks, what drives electricity prices? In Proceedings of the 11th International Conference on the European Energy Market (EEM14), Kraków, Poland, 28–30 May 2014; doi:10.1109/EEM.2014.6861289.
7. 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.

8. Monteiro, C.; Fernandez-Jimenez, L.A.; Ramirez-Rosado, I.J. Explanatory information analysis for day-ahead price forecasting in the Iberian electricity market. *Energies* **2015**, *8*, 10464–10486.
9. Ziel, F.; Steinert, R.; Husmann, S. Efficient modeling and forecasting of electricity spot prices. *Energy Econ.* **2015**, *47*, 89–111.
10. Dudek, G. Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *Int. J. Forecast.* **2016**, *32*, 1057–1060.
11. 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.
12. Karakatsani, N.; Bunn, D. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *Int. J. Forecast.* **2008**, *24*, 764–785.
13. Misiorek, A. Short-term forecasting of electricity prices: Do we need a different model for each hour? *Medium Econom. Toepass.* **2008**, *16*, 8–13.
14. Amjady, N.; Keynia, F. Electricity market price spike analysis by a hybrid data model and feature selection technique. *Electr. Power Syst. Res.* **2010**, *80*, 318–327.
15. 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.
16. 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; doi:10.1109/PEDG.2013.6785651.
17. 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.
18. Ziel, F. Forecasting Electricity Spot Prices Using LASSO: On Capturing the Autoregressive Intraday Structure. *IEEE Trans. Power Syst.* **2016**, doi:10.1109/TPWRS.2016.2521545.
19. 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*, doi:10.2202/1558-3708.1362.
20. Weron, R. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*; John Wiley & Sons: Chichester, UK, 2006.
21. Weron, R.; Misiorek, A. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *Int. J. Forecast.* **2008**, *24*, 744–763.
22. Serinaldi, F. Distributional modeling and short-term forecasting of electricity prices by Generalized Additive Models for Location, Scale and Shape. *Energy Econ.* **2011**, *33*, 1216–1226.
23. Kristiansen, T. Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy* **2012**, *49*, 328–332.
24. 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.
25. 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.
26. Maciejowska, K.; Nowotarski, J.; Weron, R. Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. *Int. J. Forecast.* **2016**, *32*, 957–965.
27. Nowotarski, J.; Weron, R. To combine or not to combine? Recent trends in electricity price forecasting. *ARGO* **2016**, *9*, 7–14.
28. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecast.* **2016**, *32*, 896–913.
29. Maciejowska, K.; Nowotarski, J. A hybrid model for GEFCom2014 probabilistic electricity price forecasting. *Int. J. Forecast.* **2016**, *32*, 1051–1056.
30. 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.
31. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263.
32. Garcia-Martos, C.; Conejo, A. Price forecasting techniques in power systems. In *Wiley Encyclopedia of Electrical and Electronics Engineering*; Wiley: Chichester, UK, 2013; pp. 1–23, doi:10.1002/047134608X.W8188.
33. Burger, M.; Graeber, B.; Schindlmayr, G. *Managing Energy Risk: An Integrated View on Power and Other Energy Markets*; Wiley: Chichester, UK, 2007.

34. Nogales, F.J.; Contreras, J.; Conejo, A.J.; Espinola, R. Forecasting next-day electricity prices by time series models. *IEEE Trans. Power Syst.* **2002**, *17*, 342–348.
35. Conejo, A.J.; Contreras, J.; Espínola, R.; Plazas, M.A. Forecasting electricity prices for a day-ahead pool-based electric energy market. *Int. J. Forecast.* **2005**, *21*, 435–462.
36. Paraschiv, F.; Fleten, S.E.; Schürle, M. A spot-forward model for electricity prices with regime shifts. *Energy Econ.* **2015**, *47*, 142–153.
37. Broszkiewicz-Suwaj, E.; Makagon, A.; Weron, R.; Wyłomańska, A. On detecting and modeling periodic correlation in financial data. *Physica A* **2004**, *336*, 196–205.
38. Bosco, B.; Parisio, L.; Pelagatti, M. Deregulated wholesale electricity prices in Italy: An empirical analysis. *Int. Adv. Econ. Res.* **2007**, *13*, 415–432.
39. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*; Springer: New York, NY, USA, 2013.
40. Bessec, M.; Fouquau, J.; Meritet, S. Forecasting electricity spot prices using time-series models with a double temporal segmentation. *Appl. Econ.* **2016**, *48*, 361–378.
41. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67.
42. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. B* **1996**, *58*, 267–288.
43. Ziel, F. Iteratively reweighted adaptive lasso for conditional heteroscedastic time series with applications to AR-ARCH type processes. *Comput. Stat. Data Anal.* **2016**, *100*, 773–793.
44. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; CRC Press: Philadelphia, PA, USA, 2015.
45. Zou, H.; Hastie, T. Regularization and Variable Selection via the Elastic Nets. *J. Royal Stat. Soc. B* **2015**, *67*, 301–320.
46. Bordignon, S.; Bunn, D.W.; Lisi, F.; Nan, F. Combining day-ahead forecasts for British electricity prices. *Energy Econ.* **2013**, *35*, 88–103.



© 2016 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/>).