LUND UNIVERSITY

SCHOOL OF ECONOMICS AND MANAGEMENT

Forecasting Swedish FCR-D Prices using Penalized Multivariate Time Series Techniques

M.Sc. Data Analytics and Business Economics

Sebastian Brugger

Lennart Wunderlich

Professor: Joakim Westerlund

Supervisor LUSEM: Luca Margaritella

Supervisor Modity: Erik Östberg

24/05/2023

Abstract

The Swedish energy market is becoming more and more sustainable, with an increasing volume and number of diversified energy sources being continuously added to the mix. To stabilize the grid frequency, auctions are held to offer energy providers incentives to produce or consume energy on short notice. This paper is applying different multivariate time series models and their ensemble to find reliable forecasts. Given the high dimensionality through a variety of factors influencing the energy market, penalized models are used to perform variable selection and obtain sparser models. The investigated data contains a lot of noise. Therefore, part of the work focuses on the effect of noise filtering. The goal is to create reliable price forecasts which may help sustainable energy providers maintain their position or enter this market, stabilize the grid, and help Sweden make the transition to renewable energy.

Contents

List of Figures

List of Tables

Acronyms

A1 First FCR-D auction that takes place at D-2, two days ahead of the delivery day.

A2 Second FCR-D auction that takes place at D-1, six hours ahead of the delivery day.

AdaLasso Least absolute shrinkage and selection operator.

aFRR Automatic frequency recovery reserve.

AIC Akaike information criterion.

AR Autoregressive.

ARIMA Autoregressive integrated moving average.

ARMA Autoregressive moving average.

FCR-D Frequency containment reserve - disturbance.

FCR-N Frequency containment reserve - neutral.

KPI Key performance indicator.

Lasso Least absolute shrinkage and selection operator.

MA Moving average.

MAE Mean absolute error.

MSFE Mean-squared forecast error.

NN Neural network.

OLS Ordinary least squares.

SES Simple/single exponential smoother.

SVK Svenska kraftnät.

TSO Transmission system operator.

VAR Vector autoregressive.

1 Introduction

Through the increase of renewable energies in the Swedish energy market, the frequency of the energy grid is under constant threat of deviating from the nominal frequency of 50Hz. Solar and wind power are weather dependent and thus behave randomly, and regulation in case of frequency disturbance is becoming ever more important. Since January 2022, Svenska kraftnät (SVK) is therefore providing several auctions that are offered to licensed Swedish energy providers like Modity AB. These companies can place bids for hourly prices and volume for either providing energy on a product called Frequency Containment Reserve - Disturbance (FCR-D) Up if the frequency is lower than allowed or consuming/stopping producing energy if the frequency in the grid is too high (FCR-D Down). In case of a deviation from the grid frequency, the energy providers have to act within seconds and produce more electricity or stop production. In either case, the energy providers are paid if they win in the first auction [1], which takes place two days ahead the delivery day, or in the second auction [2], which closes six hours before the delivery day. The auction design follows the pay-as-bid approach meaning that winning entities are being paid the amount they placed in the auction for providing the service. The purpose of this paper is to provide reliable forecasting methods for Modity AB to consider in their bidding process on FCR-D Down and FCR-D Up markets, with the former being the focus of the analysis. Having such predictions can generally help renewable energy sources to become more attractive to energy providers and help accompany the transition to those resources.

Several works have dealt with forecasting applications for balancing services on Nordic markets, which will be used for identifying relevant variables and models. Given the noisy characteristics of the prices, filtering of the input variables is considered and supported by the investigated research. While the existing literature seems to focus on Neural Network (NN) approaches, this thesis tackles the problem with a classical time-series approach. The variables of interest are fitted into an Automated Autoregressive Integrated Moving Average (Auto-ARIMA) model, which is then used to forecast the prices. That model serves as a starting point. It is expected that including further variables and using a Vector Autoregressive (VAR) model can help to increase the forecasting performance further. Given a large number of included coefficients due to the lag structure of the VAR, penalized methods dealing with the Lasso regularization are considered and tested in several variants. Further improvements might then be achieved by combining the models into an ensemble forecast. The models used in this work are all time-series models and build up on each other so that a continuous improvement of the models can be seen and will provide an outlook of what could be done in future research.

[1] The first auction closes at 3 pm and is also called (A1) auction or D-2 auction.

[2] The second auction takes place at 8 pm the day after the first auction and is also called A2 auction or D-1 auction.

The performance of the models is measured against a naive predictor, which is using the last-seen observation through the forecasting period. Our out-of-sample analysis comprises $8,800$ hourly forecasts (370 days) between March 2022 and March 2023. During that period, a total of 40 model variants are compared for both FCR-D Down and FCR-D Up markets. As comparative measures, the Mean Squared Forecast Error (MSFE) and Mean Absolute Error (MAE) are the key performance indicators (KPI). Concluding the research objective, this master thesis aims at developing penalized multivariate time-series models to obtain reliable FCR-D forecasts for the Swedish market, mainly focusing on the FCR-D Down product.

The structure of the thesis is organized as follows. Section Two contains the literature review. The following chapter covers the theoretical background of the Swedish FCR-D market in more detail, as well as the theoretical background of the noise filtering methods and time series models used in this work. In Section Four, the methodology is presented. Section Five contains the analysis and is followed by a further research outlook. Finally, Chapter Seven entails the conclusion and provides practical implications of the results.

2 Literature Review

To provide a motivation for the methodology and to provide an overview of the current state of academic research in the scope of this work, in this chapter, a literature review is conducted. Since a large body of literature is available in the area of ancillary energy price forecasting, this literature review focuses on the contributions most relevant to this work. The relevance is determined by the applied methods and the geographical market focus. This chapter generally follows the structure of this work. First, research from the narrow scope of ancillary energy service price forecasting using statistical methods instead of numerical models (see for example Siddiqui et al., 2001 or Gilmore et al., 2022) is presented. To build on those results later, emphasis is put on variable selection. Then, a brief review of noise filtering for energy market data is provided followed by research on (Lasso) VAR applications in the energy market context.

The work by Pihl (2019) is closely related to the scope of this work and investigating the behavior of Swedish FCR-D and FCR-N prices between January 2016 and April 2019. In addition to the price data, data on hydro production, energy spot price, reservoir levels, aFRR[3] and time-fixed effects are used to explain the variation in FCR prices. Using all covariates, 48% of the price variance can be explained using linear regression, with the reservoir data being most important for explaining FCR-D prices. With regard to price predicting, the author finds a simple Neural Network (NN) with one hidden layer and 15 hidden units to significantly improve the forecasting performance compared to the linear regression model, especially for extreme price values. While the article can serve as an indicator for detecting explanatory variables related to FCR-D prices, it does not consider time-series methods for prediction. Furthermore, it should be noted that the data was obtained from an old market structure of the FCR-D market when there was no separation in the auction process between FCR-D Down and FCR-D Up. The results might, therefore, not be fully transferable to the new market design as explained in the next chapter but can be used as a first indication to build upon.

Giovanelli et al. (2018) investigate the predictability of the Finnish FCR-N market using NN models with a simple ARIMA(1,1,1) and Support Vector Regression (SVR) model as benchmarks. A total of 64 variables from 2015 and 2016 are collected from different categories such as electricity import/export, load, and generation as well as energy spot prices, oil prices, weather, and calendar data. They find the three-layer NN to outperform the benchmark models at most times and the ARIMA model to provide the least accurate results, especially during volatile periods. However, no tuning on the time-series model was performed, and with the ARIMA, only a univariate time-series model has been considered. Kraft et al. (2020), in contrast, consider a SARIMAX model with seven

[3]The automatic Frequency Recovery Reserve (aFRR) is an ancillary energy service aiming at restoring the frequency to the nominal value of 50Hz.

exogenous regressors in their FCR price forecasting study for continental Europe between 2014 and 2018. They conclude that while simple NN setups with one hidden layer yield the best forecasting performance, that approach lacks the interpretability of relevant FCR price predictors. Their SARIMAX model manages to beat a naive forecaster in terms of predictive performance and identifies the lagged FCR prices, future energy prices, load, and planned unavailable capacity to be the most important covariates in driving FCR prices. Again, these results might not be transferable to the current Swedish FCR-D market structure but yield interesting insights about potential covariates and indicate the potential relevance of multivariate time-series approaches.

Another forecasting research contribution with regard to the German market comes from Narajewski (2022). Using a total of 947 regressors with various lags from the load, wind and solar generation, price data (energy spot, aFRR/mFRR[4], coal, gas, and oil), and time dummies, the author addresses the forecasting of the Imbalance Price which is derived from aFRR and mFRR market data, among others, between 2018 and 2021. A regular Lasso model shows competitive results compared with NNs and Generalized Additive Models for Location, Scale, and Shape but fails to beat the naive forecast. However, a combination of the considered methods manages to beat the latter. The results provide a first indication for applying the Lasso in the context of ancillary energy service prices but do not consider the method in combination with time-series models. A further application of Generalized Additive Models is published by Hameed et al. (2023), who are predicting week-ahead FCR-N prices for several Nordic markets (Denmark, Finland, and Norway). They conclude that FCR-N products are harder to predict compared to spot markets, and smoothing curves differ for each country, even though the market characteristics are rather similar. Therefore, many inter-country market behaviors might not be present for the markets under investigation. While the research in the field of European ancillary energy service price forecasting seems to focus on the importance of NN, time-series models are rather used as benchmarks. Especially multivariate time-series analysis seems to be lacking research despite its benefits in identifying price drivers and the opportunity to combine it with variable selection methods such as Lasso as applied by Narajewski (2022). The considerable body of literature on the topic comes with the advantage to ease the identification of relevant variables to explain FCR-D prices, as explained later in this work.

Issues with the predictability of FCR prices might be related to the noisy structure of the data containing many outliers (Hameed et al., 2023; Pihl, 2019). That characteristic is likely to limit the forecasting accuracy of statistical time-series models. Therefore, several researchers argue for noise reduction when working with energy-related data (Hurta et al., 2022; Lago et al., 2021; Weron, 2014). Consequently, many different filtering techniques

[4] mFRR stands for manual Frequency Restoration Reserve and, in contrast to the aFRR, is usually manually activated by the Transmission System Operator (TSO).

have been applied. Janczura et al. (2013), for example, show that the performance of their Markov Switching Chain models for predicting Australian and European electricity spot prices can be improved when filtering the data before training the model. Several filters, such as the recursive filter on prices (RFP), are considered with none of them significantly outperforming the others. However, all of them manage to yield superior performance compared to non-filtered series. Narajewski and Ziel (2020) apply several transformations to intraday electricity price data to boost predictive performance. While Hanzák (2011) argues in favor of simple exponential smoothing (SES) to improve time-series forecasts through a GARCH approach, Hao et al. (2022) propose a more complex approach via Kalman filtering to denoise wind data for wind speed forecasting.

All these contributions reflect the heterogeneity in filtering approaches used in the academic literature. Hurta et al. (2022) notice that not only the filter selection but also its parameter choice determine the improvement in accuracy through its application. With their investigation of the impact of several filtering methods on electricity price forecasting via a Seasonal Component Autoregressive with Exogenous Factors (SCARX) model, Afanasyev and Fedorova (2019) aim at closing that gap in the existing literature. In line with Janczura et al. (2013), they find that filtering of the original data, in most cases, increases the forecasting accuracy and an ensemble of different filtering techniques, such as threshold and standard deviation filters (SFP) on prices, yield competitive results. That result is confirmed by Shah et al. (2021) for data on Italian electricity spot prices. They show that a combination of the RFP or Moving Window Filter on Prices (MFP)[5], together with threshold value replacement, help to decrease forecast inaccuracy. Further, they conclude that a multivariate Vector Autoregressive (VAR) model in all investigated cases outperforms univariate time-series models such as AR and ARIMA which leads to the next focus of this literature review.

Several authors have applied VAR models to exploit multivariate data for the prediction of energy-related data. García-Ascanio and Maté (2010) show that using a VAR model in forecasting Spanish energy demand between 2006 and 2007 with interval time series leads to more accurate results compared to multi-layer NNs. A more recent study investigates energy demand forecasts generated by VAR and VARX (a VAR model with exogenous variables) for France between 2010 and 2018 and concludes the method to be appropriate in the short- and long-run (Auray and Caponi, 2020). For Nordic energy spot markets, Haldrup et al. (2010) show that simple VAR models, in some cases, can improve the forecast accuracy when compared to a univariate RS-SARFIMA (Regime Switching Seasonal Autoregressive Fractionally Integrated Moving Average), especially when considering regime-switching VAR models.

Given the high dimensionality of VAR models, they are often combined with penalization techniques to obtain sparser model representations (Dowell and Pinson, 2016; He

[5]The MFP is a modification of the SFP with a fixed rolling window.

et al., 2015). A natural extension of the VAR model to perform variable selection in the presence of many coefficients is the Lasso (Least Absolute Shrinkage and Selection Operator, Tibshirani, 1996). Cavalcante et al. (2017) provide a comprehensive review of different Lasso VAR approaches, including the standard and row-wise Lasso VAR, and apply those models to wind power plant data to forecast their production. They conclude that none of the structures serves as a universally best solution, and different approaches might maximize the forecasting ability for different data structures. However, all variants produced better results than a regular VAR and an AR model. The outperformance compared to the AR model seems to decrease with an increasing forecast horizon. Nicholson et al. (2020) confirm the outperformance of Lasso VAR compared to the regular VAR and AR model applied to stock market data and energy use data for appliances. Additionally, the Lasso sometimes outperforms their proposed HLag (Hierarchical Lag Structures) model with embedded lag selection. Further evidence for the relevance of the Lasso VAR procedure comes from Messner and Pinson (2019). They show that for wind power forecasting in Danish and French markets the Lasso VAR consistently outperforms a time-adaptive AR model. Additionally, they expand the Lasso VAR approach with a time-adaptive component exerting exponentially decaying weights on data from the past. This approach allows the model to adapt to changes in the data dependencies, supports efficient computing, and outperforms the regular Lasso VAR in the exercised forecasting application.

Concluding the literature review, it is observed that the research on ancillary energy markets, especially on FCR markets, to date seems to focus on investigating the forecasting abilities of NNs instead of multi-variate time-series applications. So far, only exogenous variables have been added to an ARIMA model setup, as in Kraft et al. (2020). Given that FCR prices are usually characterized by noisy data structures, noise reduction through data filtering might be appropriate. However, there seems to be no single-best approach to apply to energy-related data and several methods have been proposed by the academic literature depending on different data characteristics. Connecting to the lack of multivariate time-series applications in the ancillary energy price forecasting research, the VAR model seems to be an appropriate choice that has been successfully applied to other data from the electricity market topic. A natural extension of the VAR model, given its often large number of coefficients, is the Lasso regularization. Promising research has been published underlining the potential of sparse Lasso VAR models to outperform the standard VAR and univariate time-series models.

3 Theoretical Background

This Section covers the theoretical background of the investigated market and the methodologies described in the next chapter. Since the data used here exhibit a lot of noise, noise filtering methods are considered to pre-process the data. The Auto-ARIMA model is introduced as a baseline model, followed by the Vector Autoregression (VAR) framework, which allows for multivariate time series analysis. Given the high dimensional structure of the analyzed data, a Lasso VAR procedure is introduced to obtain a sparser forecasting model. To increase the computational efficiency, an adaptive version of that approach as described by Messner and Pinson (2019) is considered and briefly explained in the final part of the Chapter.

3.1 FCR-D Market in Sweden

The FCR-D market emerged in 2018 and underwent a significant change in 2022. Before that year, the FCR-D market represented a single market. Now, the market is divided based on frequency ranges, either above the allowed 50Hz grid frequency, or the range below. 50% of the service needs to be provided within 5 seconds, and the full service needs to be activated within 30 seconds. The service should endure for up to 30 minutes (Kraftnät, 2023a). Since the new market introduction, the market volume has been increased stepwise (usually quarterly) with a target market volume of 558 MW (FCR-D Up) and 538 MW (FCR-D Down)(Kraftnät, 2023b).

Before 2022, energy providers were required to bid for both frequency ranges at the same time, which restricted the potential sources of energy. Notably, hydro plants dominated the FCR-D market due to their ability to regulate their energy production at will. Modity AB provides bidding services with a particular focus on the FCR-D Down market. The FCR-D Down market is especially relevant for wind farms, as they have the ability to reduce their energy output but cannot increase it in the absence of wind. This is precisely why the market was split into two frequency ranges. Without this split, wind energy providers would be unable to participate, as they would be required to increase their production to meet increased demand, even when they lacked sufficient wind to do so. Not being able to meet the required demand leads to heavy penalty payments. The FCR-D market offers significant advantages to sustainable energy companies, enabling them to operate at maximum capacity and adjust their energy production as needed to generate additional revenue.

As previously noted, the auction process occurs in two stages. The first stage begins seven days prior to the delivery day, where companies are allowed to submit their bids until 3 pm, two days before delivery day. During this stage, companies may specify the prices and volume they can provide, whether in production or reduction/consumption, for each hour on the delivery day. The lowest price is ultimately selected, and all lowest prices are accepted until the required volume for A1 is met. All accepted bids are then averaged, and only the resulting averaged price is published the latest two hours after the end of the corresponding auction. Figure 1 illustrates the available average prices at the closing of the gate for A1. As most of the variables are sourced from mimer.se or are forecasts, they are available even beyond the gate closing time. Specifically, A1 data from D-3 is available until the end of D-1, thereby enabling predictions to be made from that point forward. However, certain variables, such as A2 data, then cannot be incorporated into the multivariate time series models utilized to predict A1 prices. There exists a trade-off between the incorporation of more variables and the necessity of forecasting further into the future versus utilizing fewer actual variables and simulating closer proximity to the delivery day.

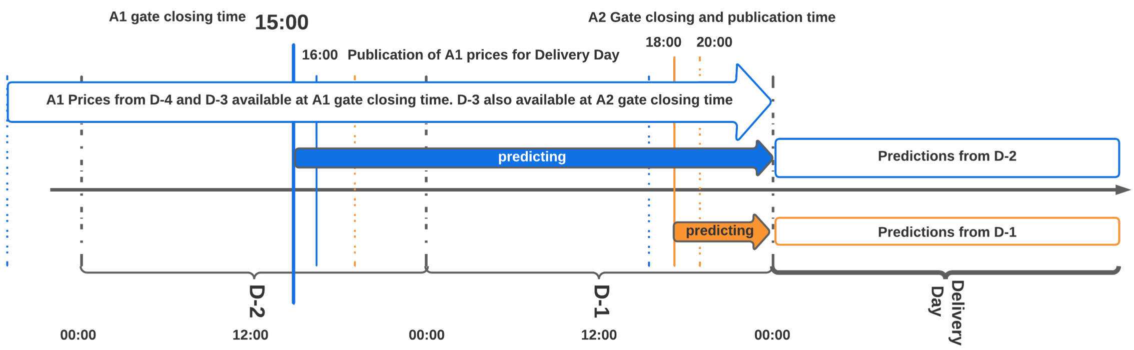

Figure 1: The timeline highlights important times for the first auction A1. The bold blue vertical line at 15:00h shows the gate closing time, while the dashed line an hour later indicates the results of the auction. The blue colors represent data or related points to A1 prices, while the orange represents elements of A2, and its dedicated schedule can be found in Appendix 9.

3.2 Noise Filtering

The noise filtering methods explained here are considered as a data pre-processing step. Therefore, it is refrained from giving a detailed introduction of the methods, and only an intuitive introduction of the applied methods is provided. Three methods are described (in order of complexity): Frequency Averaging, Exponential Smoothing, and Kalman Filter[6].

The underlying data consists of hourly measurements collected over the course of a year, yielding roughly 11,000 data points. To reduce the number of observations, frequency averaging is employed to average sequential data that are in close proximity. In this research, a moving average is utilized to merge two-hour blocks of data, resulting only in a slight noise reduction (Enders, 2015). While a stronger approach may be feasible, it risks oversimplification and may hinder the prediction of spikes. Frequency averaging, however, provides more than just noise reduction benefits. Compressing the

[6]It should be noted that there are more applicable filtering techniques available. See Hyndman and Athanasopoulos (2021, Chapter 7) for a more comprehensive overview.

data enables a compressed forecasting horizon, meaning fewer steps into the future need to be predicted (Hyndman and Athanasopoulos, 2021).

The exponential smoother works recursively and is simple to apply to real-time data. Despite its easy implementation, it yields competitive results compared to rather complex techniques. Examples of frequent applications are stock market data and inventory control (Orfanidis, 2018). While many exponential smoothing variants exist, the simple/single exponential smoother (SES) is used here. The filter assigns larger weights to recent observations compared to long-past observations. That is obtained by calculating a weighted average with exponentially decreasing weights. The methodology is described in the equation below, where $0 \leq \alpha \leq 1$ acts as the smoothing parameter (Hyndman and Athanasopoulos, 2021). Smaller values for α imply a smoother filtering of the series, while larger α values result in a filtered series being closer to the original one.

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T l_0, \tag{1}$$

where $\hat{y}_{T+1|T}$ is the filtered observation of y at time $T+1$ given the observations until T and l_0 reflects the initial observation. Applying SES, therefore, requires deciding on the parameters α and l_0.

The Kalman Filter is introduced as the most complex one in the set of considered filters. It is a state space model, and as the SES recursively provides estimations or predictions based on previous observations (Kalman et al., 1960). Since its first implementation, it has been subject to extensive research, mainly in the area of dynamic systems such as target tracking and noise cancellation (Badhwar and Sappal, 2016). For the computation only the previous filtered value (state estimate) and a new observation are required, resulting in high computational efficiency due to the low memory requirement (Kamen and Su, 1999). The filtering algorithm is designed to minimize the Mean Squared Error (MSE) recursively (Haykin, 2002) and deducts parameters from noisy observations (Badhwar and Sappal, 2016). Although there are many variants of the filter available, the standard version of the filter with the simplification for noise reduction as provided by Ma'arif et al. (2019) is used. The algorithm comprises a prediction and an update step.

Prediction: [7]

$$\hat{y}_{T|T-1} = \hat{y}_{T-1|T-1}, \tag{2}$$

$$P_{T|T-1} = P_{T-1|T-1} + Q, \tag{3}$$

[7] In the implementation by Ma'arif et al. (2019), the process variance is time-variant (Q_T). In this work, a time-fixed Q is used due to simplicity.

Update:

$$K_T = P_{T|T-1}(P_{T|T-1} + R)^{-1}, \tag{4}$$

$$\hat{y}_{T|T} = \hat{y}_{T|T-1} + K_T(y_t - \hat{y}_{T|T-1}), \tag{5}$$

$$P_{T|T} = (1 - K_T)P_{T|T-1}, \tag{6}$$

where $\hat{y}$ represents the filtered observation (estimated state), P is the state estimate variance (estimate uncertainty), Q is the fixed process variance, R is the fixed measurement variance, and K is the Kalman gain. Once values for y_0 and P_0 have been initialized, the first prediction can be obtained from the prior state estimate ($\hat{y}_{T-1|T-1}$). The fixed process variance Q acts as an uncertainty parameter that will always be added to the prior state estimate variance. The Update step of the algorithm then measures the current state of the system and provides the updated and filtered observation $\hat{y}_{T|T}$. The new value depends on its previous estimate $\hat{y}_{T|T-1}$ and the weighted difference between the new observation y_t and $\hat{y}_{T|T-1}$. The Kalman gain, defined in Equation 4, serves as the weighting of the difference. From its definition, it is close to zero with high measurement uncertainty R and low estimate uncertainty ($P_{T|T-1}$). Intuitively, the Kalman gain would put a large weight on the difference term in Equation 5 and vice versa with low measurement uncertainty and high estimate uncertainty. Since $0 \leq K \leq 1$, Equation 6 implies a decreasing estimate uncertainty as more observations become available. Both, the measurement variance R and the process variance Q are implemented as fixed parameters in this application. Mainly the measurement variance R is tuned to vary the degree of noise reduction resulting from the approach. As larger values lead to a smaller Kalman gain, choosing high parameter values for R leads to stronger filtered series.

3.3 Auto-ARIMA

The automated-ARIMA from the forecast package in R is a powerful and easy-to-use tool that can be used to find the best fitting ARIMA model to a univariate time series (Hyndman and Khandakar, 2008). The authors state that this function is created to overcome the difficulty of finding the right order, which is often-times difficult or subjective. An ARIMA model is an integrated autoregressive and moving average (ARMA) model that in turn consists of the autoregressive (AR) and the moving average (MA) component. The AR(p) component is a parametric estimator that uses p lags of itself to predict its value at time t. See Equation 7 for an AR(2),

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t, \tag{7}$$

where a_0 is a constant, $a_1,\ a_2$ are coefficients, y_t is the univariate time series at time t and ϵ is a random error component at time t, which is assumed to be normally distributed with mean 0 and constant variance σ^2. The MA(q) component is using past errors to predict the current value of the time series. Equation 8 shows a MA(2) process and both components together can be found in Equation 9, resulting in an ARMA(2,2),

$$y_t = \mu + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + u_t, \tag{8}$$

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + u_t, \tag{9}$$

where μ is the mean of the time series, $\beta_1,\ \beta_2$ are coefficients, and u_t is an error term with the same properties as ϵ. ARMA models assume stationarity, which is required to get meaningful coefficients. Stationarity means that the time series is time-independent in terms of mean, variance, and covariance (Enders, 2015). That usually requires some pre-processing of the data. That is performed by the integrated term, which indicates the order of differencing to be made to achieve stationarity. This means an ARIMA model can process most unprocessed data. An ARIMA(p,d,q) is differenced d times and has p and q lags for the AR and MA component respectively. The *auto.arima* function selects the optimal lags p, q, as well as how often the data is differenced to find the best fit according to the Akaike's Information Criterion (AIC). Hyndman and Khandakar (2008) point out that the *auto.arima* function is mainly used to select a suitable model order, including the seasonality components (P,D,Q). p,q (and P,Q) can be selected via the AIC if d and D are known. To see how the AIC is used to choose the orders, see Hyndman and Khandakar (2008) Chapter 3.1.

3.4 Vector Autoregression (VAR)

VAR models extend the univariate AR model to a dynamic multivariate time series model. Given its flexibility, easy implementation, and often superior forecasting performance compared to univariate models, it has become extensively used in economic and financial time series applications. While the usage of the VAR model is not restricted to forecasting purposes but can also serve for structural inference and policy analysis - given certain assumptions - (Zivot and Wang, 2003), the scope of the model is restricted to the forecasting functionality in this work. The p-lag VAR(p) model can be defined as

$$\boldsymbol{y}_t = \boldsymbol{c} + \sum_{l=1}^{p} \boldsymbol{A}_l \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t, \tag{10}$$

with $\boldsymbol{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})'$ being a $(n \times 1)$ vector of n time series variables at time t, $\boldsymbol{c}$ reflecting a $(n \times 1)$ vector of constants, $\boldsymbol{A}_l$ are $(n \times n)$ coefficient matrices for each lag, and $\boldsymbol{\epsilon}_t$ being an unobservable white noise error process characterized by a time-invariant positive-semidefinite covariance matrix $\boldsymbol{\Sigma}$. Further, it is assumed that all series in $\boldsymbol{y}_t$ are stationary (Enders, 2015). Considering Equation 10, a bivariate VAR(2) model could then be written as a system of two the two equations

$$\begin{aligned} y_{1t} &= c_1 + a^1_{11} y_{1t-1} + a^1_{12} y_{2t-1} + a^2_{11} y_{1t-2} + a^2_{12} y_{2t-2} + \epsilon_{1t}, \\ y_{2t} &= c_2 + a^1_{21} y_{1t-1} + a^1_{22} y_{2t-1} + a^2_{21} y_{1t-2} + a^2_{22} y_{2t-2} + \epsilon_{2t}, \end{aligned} \tag{11}$$

where a^1_{11} reflects the first entry in the coefficient matrix $\boldsymbol{A}_1$ for the first lag and so on. Forecasts $\hat{\boldsymbol{y}}_t$ from the VAR model can be deducted from Equation 10 and denoted as

$$\hat{\boldsymbol{y}}_t = \hat{\boldsymbol{c}} + \sum_{l=1}^{p} \hat{\boldsymbol{A}}_l \boldsymbol{y}_{t-p}. \tag{12}$$

The three previous equations result in 1-step ahead forecasts which can then be used to iteratively compute multi-step ahead forecasts. Another feasible strategy is to predict direct forecasts by leading the target vector $\boldsymbol{y}_{t+h}$ by the desired h forecasting steps. Taieb et al. (2011) provide a comprehensive overview of different forecasting techniques and conclude that the iterative forecasting strategy often outperforms the direct approach. Other authors, such as Messner and Pinson (2019), instead argue that using iterative 1-step forecasts can lead to an accumulation of forecast errors and prefer direct forecasting.

The assumption of stationary input time series for the VAR(p) model requires differencing of $I(1)$ series, where a series is considered $I(1)$ if it is non-stationary in levels but becomes stationary after being differenced once. However, there is some academic literature arguing that VAR(p) models can be estimated in levels, especially in forecasting applications. Luetkepohl (2011), for example, argues that linear forecasts from a VAR model are still MSE-minimizing even if the model contains $I(1)$ components, however, with unbounded increasing forecast uncertainty as the forecast horizon increases.

The standard VAR(p) approach is estimation via OLS and therefore the coefficient matrix $\boldsymbol{A}_l$ is obtained by minimizing the SSE

$$\sum_{t=1}^{T} \|\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t\|_2^2, \tag{13}$$

with $\|\boldsymbol{x}\|_2 = (\sum_{i=1}^{N} |x_i|^2)^{1/2}$ being the L2-norm. That property allows for finding an analytical solution to the coefficient matrix $\boldsymbol{A}_l$. Taking the quadratically increasing number of coefficients $n^2 p$ into account, with a considerable number of included time series and lags, a VAR(p) model can quickly be prone to overfitting. That could then result in distorted forecasting performance, particularly when dealing with small sample sizes (Messner and Pinson, 2019).

3.5 Lasso VAR

To limit the problems that might arise from overfitting, the Lasso regularization introduced by Tibshirani (1996) is a popular choice for variable selection in regression problems. The method has been applied to VAR models in various academic publications (see, for example, Davis et al., 2016; Hsu et al., 2008; Song and Bickel, 2011) and applied in the energy market context (Cavalcante et al., 2016; Messner and Pinson, 2019). Even though the Lasso model selection consistency is restricted to certain dependence structures of the covariates, Callot et al. (2014) argue for the Lasso method in the VAR context to select relevant covariates and estimate their coefficients with high accuracy. In addition, it combines variable selection and estimation, which is beneficial, especially in high-dimensional settings where a multiple-step procedure for variable selection and estimation would result in costly computations. The Lasso regularisation extends the objective function in 13 by a penalty term,

$$\frac{1}{2}\sum_{t=1}^{T}\|\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t\|_2^2 + \lambda\sum_{l=1}^{p}\|\boldsymbol{A}_l\|_1, \tag{14}$$

with the L_1 norm $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{N}|x_i|$ and the penalty parameter $\lambda \geq 0$. The latter is determining the sparsity induced to the model by shrinking coefficients with smaller explanatory power to zero. As a result, the coefficient matrix reduces to a sparse selection reflecting covariates with the largest contribution to the prediction. It should be noted, that the penalty is not applied to the constant of the model. In this work, the approach by Callot et al., 2014 is followed, and one Lasso VAR is estimated for each time-series y_{it} in the system. This approach allows for identifying an individual solution for each of the n time series. However, it should be noted that a solution for the complete system as applied, for example, in Cavalcante et al. (2016) can be considered as well and potentially yields better results.

In contrast to the VAR estimation by OLS, there is no analytical solution to the minimum of Equation 14 due to the non-differentiable L_1 norm. Therefore, numerical methods need to be applied in order to estimate the coefficient matrix for different sets of λ. As suggested by Callot et al. (2014) and Messner and Pinson (2019), among others, this work relies on the cyclic coordinate descent method as introduced by Friedman et al. (2010). In addition, it needs to be decided on the choice of the parameter λ, which will be more elaborated on in the model selection (4.2) part.

As Zhao and Yu (2006) notice, the model selection obtained from the Lasso method is only consistent under relatively restrictive assumptions regarding the dependence structure of covariates. Since the covariates in VAR models might be highly dependent, especially due to the lag structure, Adaptive Lasso (AdaLasso) solutions, as suggested by Zou (2006) should be considered. To introduce the AdaLasso, Equation 14 is expanded by the weights vector $\hat{\boldsymbol{w}}_l$,

$$\frac{1}{2}\sum_{t=1}^{T}\|\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t\|_2^2 + \lambda\sum_{l=1}^{p}\hat{\boldsymbol{w}}_l\|\boldsymbol{A}_l\|_1, \tag{15}$$

with $\hat{\boldsymbol{w}}_l = \|\hat{\boldsymbol{\beta}}_I\|^{-\gamma}$, $\gamma > 0$, being the set of weights for lag l, and $\hat{\boldsymbol{\beta}}_I$ being an initial $\sqrt{T}$-consistent estimator of $\boldsymbol{\beta}^*$. Following Callot and Kock (2014), the least squares estimator is used for $\hat{\boldsymbol{\beta}}_I$, and γ is chosen to be 1. The intuition of the weights is to let the AdaLasso produce more intelligent choices of the penalty term. If the true coefficient value (β^*) for any variable or its lag is zero, then its initial $\sqrt{T}$-consistent estimator $\hat{\boldsymbol{\beta}}_I$ would be close to zero as well, letting the respective $\hat{w}_{il}$ leading to a high penalty for that coefficient. Consequently, for large values of $\hat{\boldsymbol{\beta}}_I$, the penalty imposed on that coefficient would then be small. Setting $\hat{\boldsymbol{w}}_l = 1$ would result in the regular Lasso penalty. As explained in Callot and Kock (2014) this two-step procedure leads to a better asymptotic performance relative to the regular Lasso.

3.6 OnlineVAR

The idea of Lasso VAR can be further developed. In larger data sets with a large number of observations from the distant past and many frequently updated variables, it may be useful to consider only a subset of the data. In other words, only a few recent observations are considered, discarding older ones. This online component assumes that after a new observation is introduced, the parameters will change only slightly, and uses the parameters from the previous iteration as a base to compare how much has changed. Implementing this process can lead to a substantial reduction in computational time, which is valuable considering that data from the distant past may not be of high relevance for the forecasting task at hand. However, it is essential to have a sufficient amount of data available to obtain reliable coefficients that accurately represent the underlying relationships in the model. Striking a balance between selecting an appropriate data window and ensuring an adequate amount of data is crucial for obtaining robust coefficient estimates. OnlineVAR combines these two aspects: it uses the learned coefficients and goes through the data like a rolling window, emphasizing more recent data points and updating the coefficients when necessary (Messner and Pinson, 2019). The objective function of the OnlineVAR is only a slight variation of 14, where an exponential forgetting parameter ν is introduced.

$$\frac{1}{2}\sum_{t=1}^{T}\nu^{T-t}\|\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t\|_2^2 + \lambda\sum_{l=1}^{p}\|\boldsymbol{A}_l\|_1, \tag{16}$$

Since $\nu \in (0, 1)$, the closer t is to T, the greater the impact of ν^{T-t}. Their OnlineVAR also relies on the cyclic coordinate descent by Friedman et al., 2010 and is shown in Equation 17 (c.f., Messner and Pinson, 2019),

$$A_p[i, j] \leftarrow \frac{S(\sum_{t=1}^{T} \nu^{T-t} y_{t-p}[j](y_t[i] - \hat{y}_t^{(j,p)}[i]), \lambda)}{\sum_{t=1}^{T} \nu^{T-t} y_{t-p}[j]^2}, \tag{17}$$

where S is a soft thresholding operator, $\hat{y}_t^{(j,p)}[i]$ is the fitted value for $y_t[i]$ (11), and ν is the exponential forgetting parameter (Messner and Pinson, 2019). As mentioned earlier, for every new observation, a new calculation to get the coefficients would have to be made. To increase calculation speed, the learned coefficients are used as initial values for the new observations, and since the coefficients are only expected to change slightly, converge quicker. To gain a more detailed understanding of the topic at hand, it is referred to Chapter 2.2 of Messner and Pinson (2019). In this chapter, they provide in-depth explanations and insights into their model. Specifically, equations 11 to 16 in their paper offer mathematical formulations that are highly relevant to the discussed concepts.

4 Methodology

This part aims at explaining the steps applied in the methodology of the empirical analysis whose results are presented in the following chapter. The examined data set is initially described, with primary emphasis placed on the dependent variables, namely FCR-D prices. The data processing stage primarily involves the application of noise filtering methods discussed in the preceding chapter. Additionally, the characteristics of the processed data in terms of stationarity are explored and analyzed. Following, the model selection is described with choosing parameters for each time-series model being at the center of that part. The subsequent forecasting algorithm part defines the implementation of the models in an out-of-sample backtest. Lastly, the section on performance evaluation outlines the MSFE and MAE used to assess and measure the effectiveness of the generated forecasts. Notably, no detailed in-sample analysis is conducted as this work focuses on investigating the forecasting performance of the described models.

4.1 Data Description and Processing

The empirical analysis is conducted on 10,893 hourly observations starting in January 2022 until March 2023 using a total of 17 variables. The variables of interest are four price series: FCR-D Down and Up prices from their two auctions A1 (two days ahead) and A2 (one day ahead). Following the description in Chapter 3.1, the particular focus is on the FCR-D Down prices. Results for the FCR-D Up prices are included in the analysis as well. All four time series are displayed in Figure 2. They exhibit many sudden spikes,

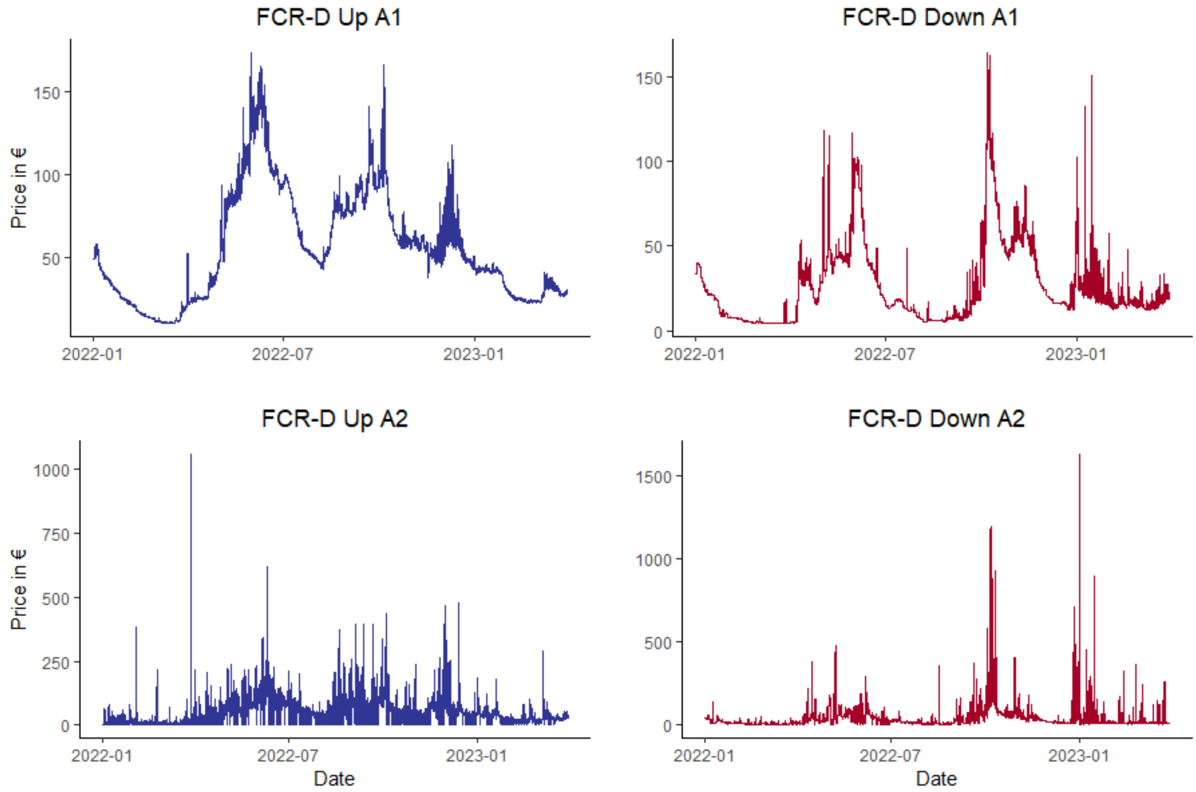

Figure 2: The four price time series of A1 and A2 auctions, with FCR-D Down (red) and FCR-D Up (blue) prices in EUR.

especially the data for the second auction, making it challenging to identify which of them are signals and which of them can be considered noise. At first glance, it seems that there might be issues with stationarity. Additionally, the within-day price structure for the FCR-D Down A1 price over time is investigated in Figure 3. With the black lines indicating market volume increases as described in Chapter 3.1 it appears that market prices also increase for a while and then stabilize at lower levels again. Generally, the market becomes much more volatile after volume increases. An additional observation, consistent with the findings of Pihl (2019), is that the prices exhibit a tendency to be higher during nighttime. This can be observed in the warmer colors at the top or bottom of 3, where the horizontal lines indicate the beginning and end of the night.

For the A2 auction, it becomes much more challenging to identify patterns as reflected in Figure 4. The prices spike throughout the year, with no clear intra-day pattern. This could be due to the more unpredictable structure of the auction. When the TSO buys enough volume in the A1 market, sometimes no or significantly less volume is requested in the A2 market. That makes the A2 market harder to predict, which is in line with the observations from Figure 2. Note, that only the plots for FCR-D Down are included as examples while the ones for both FCR-D Up auctions are part of Appendix A.

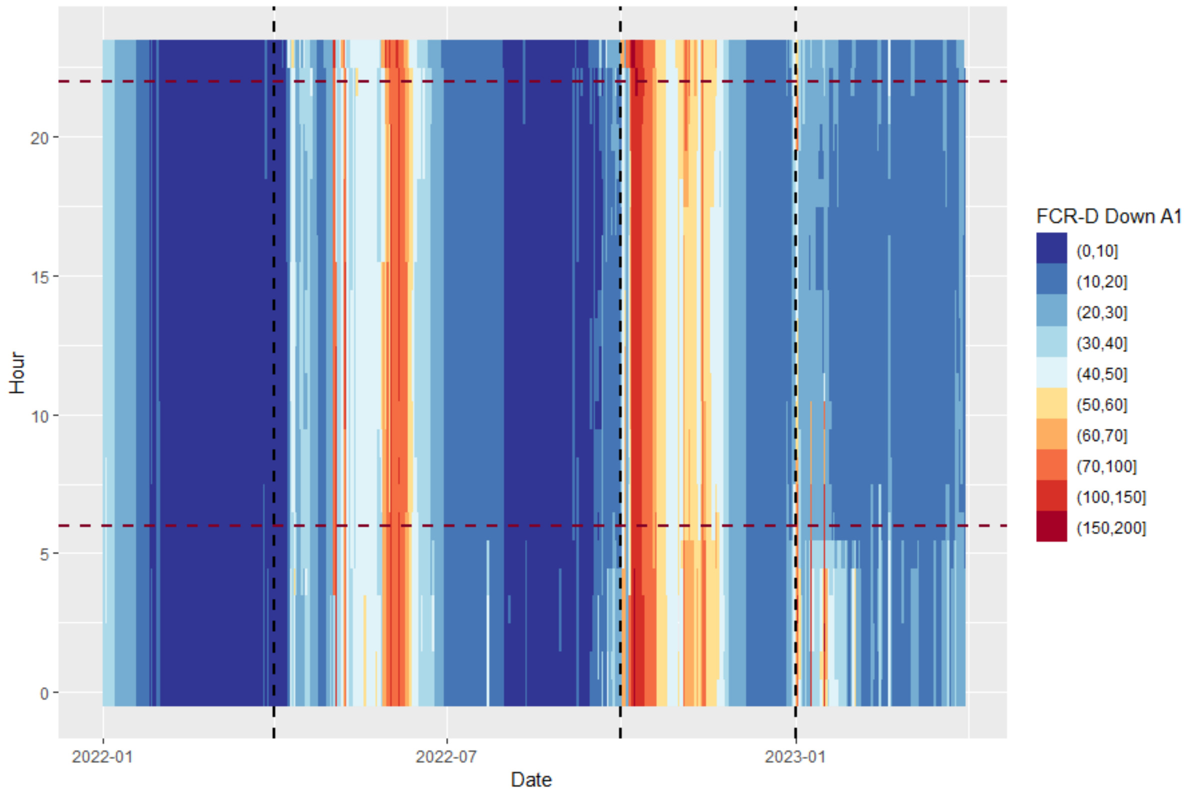

Figure 3: The heatmap shows the prices of FCR-D Down A1 for each day from January 2022 to the end of March 2023. The horizontal lines indicate the switches between day and night, while the black vertical lines indicate the beginning of a quarter when the market volume of the FCR-D is increased.

In total, 17 variables are investigated, including the price and volume data of the auctions, to better explain the variation in the FCR-D price series. A more detailed list of the variables, including their references and sources can be found in Table 7 in the Appendix as well as their descriptive statistics (Table 6). The selection of the variables is based on the findings from the literature review (Chapter 2). While Pihl (2019) argues that reservoir and hydropower production data are particularly important for explaining the variance of FCR-D prices, only the residual load is used as a proxy for the former. Unfortunately, data on hydropower production is not publicly available without being too delayed to use it in a forecasting setting. To get a better understanding of how these variables are related to the FCR-D prices, a correlation matrix is displayed in Figure 12 in the Appendix. With regard to the prices it seems that Nuclear Power Production is relatively strongly negatively correlated to the FCR-D Up A1 price, while the temperature is strongly positively correlated. Another important variable seems to be the FCR-D Up A2 price which, however, will not be used when forecasting the A1 price. This is because the A2 prices are updated after the A1 auction is closed. Including A2 data would imply to increase the prediction horizon, which seems to have a stronger negative impact on the predictability of A1 prices than including A2 prices would support the predictability. Looking at the FCR-D Down A1 price, the FCR-D Up A1 price and the FCR-D Down

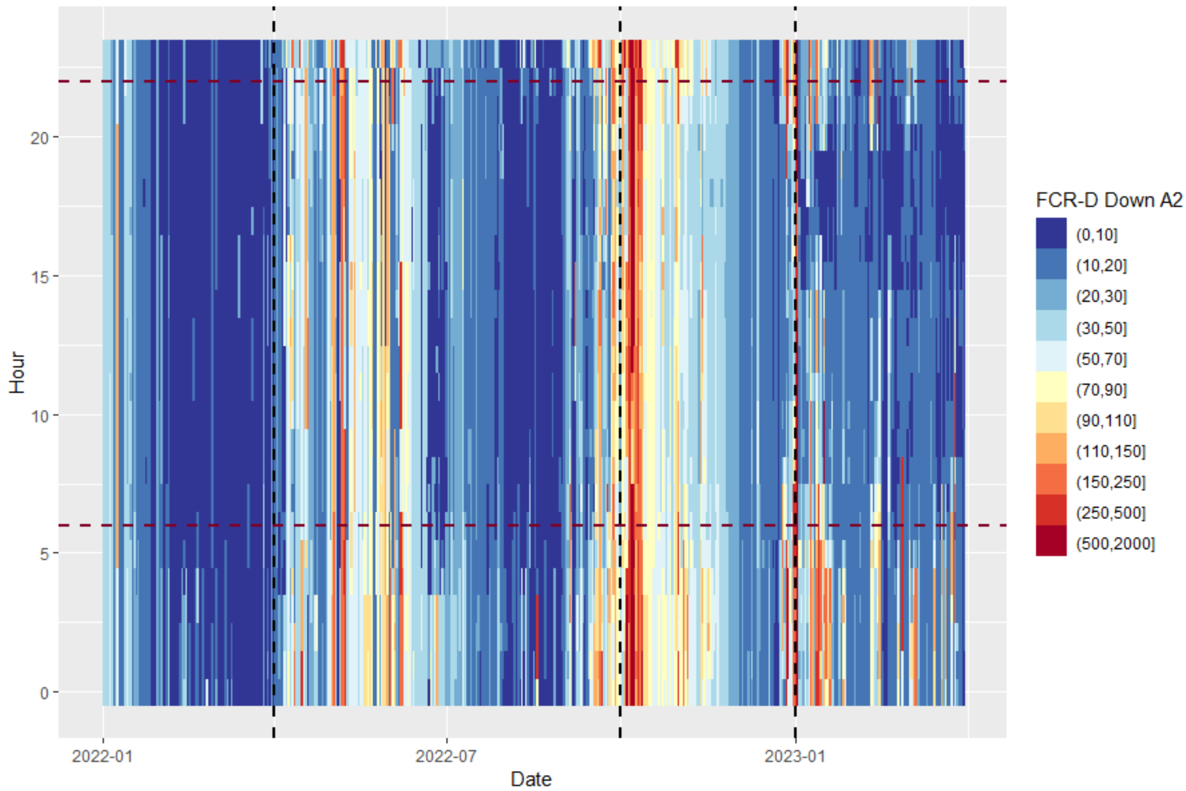

Figure 4: The heatmap shows the prices of FCR-D Down A2 for each day from January 2022 to the end of March 2023. The horizontal lines indicate the switches between day and night, while the black vertical lines indicate the beginning of a quarter when the market volume of the FCR-D is increased.

A2 price have the strongest (positive) correlations. Generally, price variables seem to be mostly correlated with FCR-D Down prices, while FCR-D Up prices seem to be correlated with production statistics and other variables such as the temperature.

The variables are checked for stationarity as this is an assumption for using the time-series models explained in the previous chapter. Most of the variables are found to be stationary using the ADF-Test (augmented Dickey-Fuller, Dickey and Fuller, 1979). For rejecting the null hypothesis of a unit root, the 10% significance level has been chosen. That is in line with the findings by Kim and Choi (2017), who argue that higher significance levels compared to the conventional ones (1% or 5%) should be used when dealing with low-power tests such as the ADF-Test. While they imply a typical range between 20% and 40%, the 10% level is used, reflecting a compromise between the conventional and the proposed levels. Using that level of significance all time series are found to be stationary except for the FCR-D Up A1 price and the weather data. While the former becomes stationary after being differenced once, a log transformation is sufficient for the latter. The stationarity characteristics are listed in Table 1 as well. In addition, the variable availability for forecasting is summarized in Table 1. For A2, there are more variables available since the gate closing time is closer to the delivery day of this product. A2 can therefore use all of the 17 variables, while A1 can consider 12.

Table 1: Variable overview and their availability for predicting A1 auctions. Additionally, it is indicated if the data is available as an actual or forecast. Stationarity is assessed through the ADF-Test using the 10% significance level. * implies that the time series needed to be transformed by either taking the first difference (diff) or log to make it stationary.

	Used in A1	**Actuals**	**Forecast**	**Stationary**
FCR-D Market				
FCR-D Down Result Price A1	yes	yes		yes
FCR-D Up Result Price A1	yes	yes		diff*
FCR-D Down Result Price A2	no	yes		yes
FCR-D Up Result Price A2	no	yes		yes
FCR-D Down Result Volume A1	yes	yes		yes
FCR-D Up Result Volume A1	yes	yes		yes
FCR-D Down Result Volume A2	no	yes		yes
FCR-D Up Result Volume A2	no	yes		yes
aFRR Market				
aFRR Down Price aggregated	yes	yes		yes
aFRR Up Price aggregated	yes	yes		yes
aFRR Down Volume aggregated	yes	yes		yes
aFRR Up Volume aggregated	yes	yes		yes
Production				
Wind Production Forecast	yes		yes	yes
Nuclear Power Production	no		yes	yes
Other				
Temperature	yes		yes	log*
Spot Price SE2	yes		yes	yes
Residual Load	yes	yes		yes

Following the results by Afanasyev and Fedorova (2019), different filter techniques can impact the forecasting performance dissimilarly. Given the uncertain research findings regarding optimal filtering techniques and since filter optimization falls outside the scope of this work, the noise reduction part is limited to the methods explained in Section 3.2. While some research contributions in Section 2 hinted at applying the Recursive Filter on Prices (RFP), it is refrained from applying that method here. This is due to its nature of removing all spikes of a time series which might yield problematic forecasting results. While it is appropriate to reduce the noise in the data, removing all spikes will likely eliminate important signals and prevent the models from predicting them. The same logic applies to the Standard Deviation Filter on Prices (SFP) although to a smaller amount. Instead, the focus is on smoothing methods that can help reduce the noisy structure of the data. Regarding SES, Hyndman and Athanasopoulos (2021) recommend a procedure based on minimizing the sum of squared residuals/errors (SSE) to obtain optimal values for α and l_0. We limit the filtering to the methods introduced in 3.2, which are applied using different parameter choices, and their respective impact on the predictive performance of the benchmark model are compared. The filter setting contributing to the best-performing benchmark predictions is then used in the remainder of the analysis. Each time series is filtered individually. For the Kalman Filter and SES, the initial values[8] are removed from the series such that they cannot deteriorate the forecasting methods.

4.2 Model Selection

Given the variety of investigated models and noise-filtering techniques, there are many combinations and tuning options to decide on. Given computational limitations and time constraints, as explained in the previous chapter, the noise-filtering performance is investigated on the Auto-ARIMA model. The best-performing filter is then identified for each of them and used for the remainder of the analysis. With regard to data pre-processing, as explained in the previous section, the FCR-D Up Price A1 is found to be stationary only after taking the first differences. However, when conducting an in-sample analysis using the models explained in Chapter 3, it is noted that the models fail to fit the differenced time series leading to an even poorer fit when transforming back to levels (see Figure 13 in the Appendix for comparison where the issue is shown using the Auto-ARIMA model as an example). Choi and Jeong (2020) provide empirical evidence for using level data even in the case of non-stationarity. They conduct a Monte Carlo simulation and real-world macroeconomic forecasting to show that using level data oftentimes outperforms using differenced data if the purpose is forecasting only.

The five models mentioned in Chapter 3 are thus all run only with the undifferenced data sets, where only log transformation on the weather data (see Table 1) is applied. Further, the models are slightly different for A1 and A2 predictions, and the impact of filtering is investigated. To be more precise, this means that for each of the four prices, the following five models are run on a data set that contains unfiltered data, and on a data set with the filtered time series. A total of 40 models are compared.

- Auto-ARIMA model with external regressors (Auto-ARIMAX)
- VAR
- Lasso VAR
- AdaLasso VAR
- OnlineVAR

Building on the result by Narajewski (2022), an ensemble forecast is added as well, defined as the mean of the predictions from models. That combination aims at increasing confidence in the forecast and thereby resulting in more reliable results. For each of the

[8]For the Kalman filter the initial values have been set to $x_0 = 0$ and $P_0 = 1$. The initial value for the SES method has been optimized to minimize the SSE according to Hyndman and Athanasopoulos (2021) and using the *ExponentialSmoothing* function from the Statsmodel Python package (Seabold and Perktold, 2010).

VAR models, optimal lags are selected through VARSelect from the *vars* package in R (Pfaff, 2008) according to the BIC (Equation 18), which usually results in a sparser model. The optimal number of lags varies between both data sets and whether it is trained on A1 or A2. The selected lag size for each setup is fed into all types of VAR models.

$$BIC_n = log(RSS_n) + \frac{log(T)}{T} nK^2, \tag{18}$$

where RSS_n is the sum of squared residuals from the trained VAR(p) model, n is the lag order, and K refers to the total number of endogenous variables (Pfaff, 2008).

When specifying the Lasso VAR models (Chapter 3.5), the λ parameter needs to be chosen. While it is common in regular Lasso regression analysis to choose the optimal λ through cross-validation (CV), it is selected through the BIC (Equation 19) in this paper. That is in line with Hecq et al. (2021), showing the superiority of BIC in selecting the Lasso penalty parameter using Monte Carlo simulations and the considerations in Kock and Callot (2015). Additionally, Bergmeir and Benítez (2012) show that standard cross-validation approaches might not be appropriate for time-series data. The criterion in that scenario is formally described as in Equation 18, modified with the new purpose of selecting the optimal λ

$$BIC_{\lambda_T} = log(RSS_{\lambda_T}) + \frac{log(T)}{T} df(\lambda_T), \tag{19}$$

with RSS_{λ_T} reflecting the residual sum of squares depending on a certain choice of λ_T. $df(\lambda_T)$ represents the model's degrees of freedom (number of non-zero coefficients; Bühlmann and van de Geer, 2011; Friedman et al., 2010) given λ_T. For each series, the BIC_{λ_T}-minimizing λ_T is chosen. The lambda grid is chosen as in Friedman et al. (2010). Since the number of observations for training a model is larger than the number of covariates, λ_{min} is chosen to be close to zero (0.0001). A grid of 100 λ values on a linear log-scale is then generated up to λ_{max}, which refers to the smallest λ value such that all coefficients would be shrunk to 0. For the AdaLasso, the parameter γ needs to be chosen. To prevent the risk of overfitting through a grid search and to build on the expertise of Callot and Kock (2014), the value is kept fixed at 1 in this work.

4.3 Forecasting Algorithm

Once all the methods and their tuning procedures have been specified, it is time to establish the framework for the forecasting algorithms. Consistent comparability across all models is achieved by employing the same setup for forecasting across all models. This means all models generate out-of-sample forecasts through an expanding window while having the same starting window size of three months[9]. Therefore, the out-of-sample

[9]Three months are 2016 hours.

analysis starts end of March 2022, and 8880 hours (370 days) are predicted. Forecasting follows the explanations in Chapter 3.1 for the FCR-D markets (see Table 1). For both auctions, the forecast horizon is one to 24 hours ahead to obtain predictions throughout the delivery day. That means every 24 hours the next 24 hourly values for the delivery day are predicted. Despite the gate time for delivery closing before (or two days before) the actual delivery day, there are already some available actuals and forecasts. It can therefore be simulated that the auction takes place exactly one hour before delivery day, as the necessary data is available. However, that implies some potential independent variables are sacrificed for the trade-off of having a shorter predicting horizon. Further, it should be noted that only historic forecast data for wind production is available, and therefore it needs to be assumed that for the other variables such as temperature, nuclear power production, and residual load, accurate forecasts are available.

The data utilized in the analysis consists of historically realized values, yet it is reasonable to assume that the underlying forecast data that Modity has access to closely approximates the realized values. However, the backtest will most likely result in better forecasting performance compared to real-world performance using actual forecasts. Additionally, the Auto-ARIMA model is run with exogenous variables to see if they are able to improve predictability. The packages used for the forecasts are:

- Auto-ARIMA: forecast package in R (Pfaff, 2008),
- Lasso VAR: lassovar package in R (Callot et al., 2014),
- OnlineVAR: OnlineVAR package in R (Messner and Pinson, 2019).

For the Auto-ARIMA model and the OnlineVAR, a recursive forecast is used, while for all other VAR models, a direct forecast is implemented[10]. Callot and Kock (2014) argue that the advantage of using direct forecasts is the adaption to the specific forecast horizon which could make the approach more robust at long forecast horizons. In addition, they show in their research, while applying the lassovar package, that using direct versus recursive forecasts does not yield significantly different results.

4.4 Performance Evaluation

The models' performance is evaluated based on the Mean Squared Forecasting Error (MSFE) and Mean Absolute Error (MAE). Both metrics are commonly used in the forecasting literature for energy market topics (see, for example, Messner and Pinson, 2019; Narajewski and Ziel, 2020). While the MSFE is described as the optimal measure for least squares applications, it is sensitive to outliers. The MAE, in contrast, is more robust in terms of outliers. Both performance measures are included as their combination can help to reveal performance differences with regard to outliers and are defined through

[10]It should be noted, that the OnlineVAR is done recursively due to easier implementation and shorter calculation time for conducting the backtest.

$$MSFE = \frac{1}{T}\sum_{t=1}^{T}(P_t - \hat{P}_t)^2, \tag{20}$$

with P_t being the original price at time t and $\hat{P}_t$ being the forecasted price at time t,

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|P_t - \hat{P}_t|. \tag{21}$$

To get a more meaningful interpretation of the models' results, their performance is compared against a naive forecast, using the last seen observation over the forecasting period. Additionally, the performances will be compared over the full testing period and month-wise. Using that approach will allow for drawing conclusions about the performance of the models during different market conditions and accordingly provide some recommendations about the model confidence with higher or lower market volatility.

Negative price forecasts are bottomed at 0. In addition, to prevent extreme forecasts from deteriorating their performance evaluation too heavily, especially for the MSFE, they are filtered out for the A1 price forecasts. Callot and Kock (2014) and Swanson and White (1995) argue for such a procedure, and Kock and Teräsvirta (2016) show that filter tuning for that purpose does not greatly influence the results. Here, the suggested approach by Swanson and White (1995) is followed and defines a forecast outlier as a forecast being outside three times the standard deviation of the estimation period plus/minus the last seen observation. In case a forecast is identified as an outlier, it is replaced with the last seen observation. Using that approach for A1 predictions sometimes helps to increase the performance in terms of both metrics and never yields worse results. The forecasts for A2 prices are not filtered out given the many spikes in the data. Given the noisy structure of the original data, better results can be achieved by getting closer to the original outliers and filtering often decreased the performance in terms of the KPIs.

5 Analysis

After introducing the theory and methodology, the empirical analysis is conducted in this part. The results from selecting the optimal filtering method for each of the FCR-D price variables are presented in Table 2. For all remaining variables, the results are presented in Appendix 8. To ease the readability, only the results based on the Auto-ARIMA model for the best-performing filter are reported relative to the unfiltered series. For the FCR-D prices the performance is measured in terms of the MSFE, whereas for the remaining variables, the MAE has been used. In line with the results from Chapter 4.1 where the FCR-D price data has been found to be particularly important, the MSFE as an outlier-sensitive metric is chosen for those series to make sure the filter still allows the model to fit the spikes in the data to a reasonable extent. The underlying argument is that some spikes in the filtered series might serve as signals and are therefore useful information when predicting. For the remaining variables, the purpose is to use the filter in order to remove uninformative noise and therefore evaluate the filters' performance based on the outlier-robust MAE. However, it should be noted that usually there was no large difference between the MSFE and MAE improvements.

Table 2: Noise filter performance on all variables. For each variable, the out-of-sample performance of the Auto-ARIMA using the best-performing filter setup is compared to an Auto-ARIMA using the original series. The performance values are compared in percentage based on the MSFE (for FCR-D prices) and MAE (other variables). For the optimal filter, K refers to the Kalman Filter and F to Frequency Averaging with the respective parameter choice in parentheses. For the Kalman Filter parameter values for R^2 between 0.1 and 1.5 have been considered, whereas for Frequency Averaging a frequency of 2 has been used.

Variable	**Performance in %**	**Optimal Filter**
FCR-D Down Price A1	+14.66	K (0.25)
FCR-D Up Price A1	+21.05	K (0.75)
FCR-D Down Price A2	+0.66	K (0.25)
FCR-D Up Price A2	−1.56	F (2)

Using filtered time series for producing Auto-ARIMA forecasts seems to work particularly well for the A1 prices, whereas for A2 prices it does not seem to improve the forecasts. For the FCR-D Up A2 prices, it even decreased the forecasting performance. Generally, for most variables, the filtering seems to only gradually improve the predictability, except for some of the aFRR data where it shows clearly better results. Oftentimes, less filtering yields better results. In terms of filter performance, the Kalman filter, which is the most complex filter investigated here, is selected most of the time. The frequency averaging method gets selected two times, while the SES never yields the best performance. To assess the impact of filter settings in a multivariate setting, a filtered data set was created, using the optimal filter settings determined in this study. The filtering resulted in poorer performance in six out of 17 cases. The respective series are then included in the data set for the backtest without any filtering.

Investigating both unfiltered and filtered data, the results on FCR-D Down data are presented first. The models are ordered by their complexity. Note, that the Auto-ARIMAX setup in that comparison includes exogenous regressors and is therefore not the same model used for the filtering selection. For the A1 market, the results are mostly in line with expectations. Whereas the naive forecast performs worse, using more complex models increases the forecasting performance (see Table 3). Adding more information in a time series fashion through the (penalized) VAR models helps to gain further increases in forecasting accuracy. Only the OnlineVAR seems to fall behind its penalized VAR peers, which will be elaborated further on later in this section. Combining all models (excluding the naive estimator) into an equally weighted averaged ensemble forecast yields the best performance in terms of MSFE, which is in line with the previous reasoning (see Chapter

Table 3: Out-of-Sample forecasting performance for the FCR-D Down prices. For both auctions, unfiltered and filtered data are compared. For each section and metric, the best-performing model is highlighted in bold.

FCR-D Down	**A1**				**A2**			
	Unfiltered		**Filtered**		**Unfiltered**		**Filtered**	
Model	MSFE	MAE	MSFE	MAE	MSFE	MAE	MSFE	MAE
Naive	108.79	4.52	85.79	4.44	7496.71	**29.66**	7657.10	31.28
Auto-ARIMAX	79.90	**4.01**	60.28	**4.14**	5813.18	29.93	**5812.28**	**29.94**
VAR	62.18	4.61	62.05	4.55	5818.79	35.79	5948.31	32.85
Lasso VAR	61.05	4.72	63.03	4.76	5554.27	34.96	6077.98	33.36
AdaLasso VAR	60.54	4.35	58.73	4.19	5676.69	35.29	6458.41	33.87
OnlineVAR	97.88	5.04	101.38	5.53	8988.93	31.02	9083.58	31.29
Ensemble	**54.14**	4.15	**57.83**	4.37	**5483.44**	31.26	5865.39	30.46

4.2). In terms of MAE the Auto-ARIMAX model performs best for the A1 auction but lags behind some of the other models in terms of the MSFE. That could be an indication for the model to provide generally reliable predictions but missing some of the outliers in the data more often compared to other models such as the AdaLasso VAR. Furthermore, that aspect highlights the relevance of the ensemble forecast combining models that might serve different purposes. While the Auto-ARIMAX might be better for getting stronger predictions on an absolute basis, the AdaLasso VAR, for example, might be better for predicting spikes in the data. Applying filtering to the data seems to help increase the forecasting performance for A1 on some models, whereas the ensemble forecasts are a bit worse in terms of both metrics. However, while there is a large performance gain for the Auto-ARIMAX method, the filtering only has a minor effect on the predictions. As for the unfiltered series, the best-performing models are the ensemble (MSFE) and Auto-ARIMAX (MAE) with slightly worse results. These results could hint at the elimination of important signals by the applied noise filtering techniques.

Generally, the previous assumption that the A2 market might be harder to predict is reflected in the empirical results. As mentioned before (see Chapter 4.1), the A2 market behaves differently. Here, the data contains much more spikes. In terms of the MAE, no model is able to beat the naive benchmark. Low-complexity models could have an advantage in such hard-to-predict environments. They tend to provide flatter forecasts, and, therefore, are less risky. The more complex models, in contrast, are able to predict spikes in the data as well. In the noisier A2 market data, the models fail more often and more extreme. The ensemble method seems to help balance out both effects and provides the best-performing results for the unfiltered data. Filtering the A2 data worsens the results for most models in terms of MSFE, and only sometimes results in better MAE numbers. That provides further evidence for the elimination of signals relevant information for predicting the FCR-D prices.

The FCR-D Up prices seem to be more predictable and again, all models except for the OnlineVAR manage to outperform the naive forecast. Adding more information in the

Table 4: Out-of-Sample forecasting performance for the FCR-D Up prices. For both auctions, unfiltered and filtered data are compared. For each section and metric, the best-performing model is highlighted in bold.

FCR-D Up	**A1**				**A2**			
	Unfiltered		**Filtered**		**Unfiltered**		**Filtered**	
Model	MSFE	MAE	MSFE	MAE	MSFE	MAE	MSFE	MAE
Naive	47.06	3.48	56.61	4.42	**2329.50**	**26.88**	**2336.69**	**27.06**
Auto-ARIMAX	45.76	3.68	**47.45**	**3.92**	2649.75	30.57	2634.30	30.38
VAR	41.60	3.67	52.06	4.20	2473.00	30.17	2500.70	30.56
Lasso VAR	40.56	3.47	54.02	4.30	2465.45	28.96	2458.76	29.55
AdaLasso VAR	46.16	3.57	54.47	4.35	2460.69	29.21	2533.46	30.16
OnlineVAR	57.85	4.36	77.26	5.36	2731.27	31.44	2761.26	31.78
Ensemble	**37.91**	**3.32**	53.32	4.31	2431.55	28.48	2448.07	28.88

form of additional variables to the forecasting approach seems to increase the performance using the VAR and Lasso VAR model compared to the Auto-ARIMAX approach. Using the combination of the models again yields the best performance in terms of both, MAE and MSFE. Using the filtered data set for the predictions worsens the results for each model on the A1 data and sometimes yields minor improvements for the A2 market. For the A2 market, the naive forecaster performs best, both in terms of MAE and MSFE. That highlights the difficulty of accurately forecasting the A2 market and might also indicate that using the previous day's last price obtains relevant information for predicting the next day. Generally, it should be noted that using a multivariate time series model can often help to improve the forecasting performance compared to the Auto-ARIMAX model, which incorporates exogenous variables already. However, there is no single best model choice. For example, using the VAR model without penalization can result in similar or even better performance compared to the more complex penalized models. A reason could be that the VAR model might be at the risk of overfitting given its large number of coefficients, it still might be able to catch some signals in the data that are helpful for predicting the FCR-D prices and would have been eliminated by the penalized methods. However, for Down A1 Filtered/Unfiltered (AdaLasso VAR), Down A2 Unfiltered (Lasso VAR), Up A1 Unfiltered (Lasso VAR), Up A2 Unfiltered (AdaLasso VAR), and Up A2 Filtered (Lasso VAR), one of the penalized models is the best choice in terms of MSFE among the investigated models. That means that in six of the eight investigated scenarios, a penalized VAR model is outperforming its competitors. Equally often, the ensemble method manages to improve the performance of an individual best model in terms of MSFE. In terms of MAE, Auto-ARIMAX shows competitive results, especially on the FCR-D Down markets.

The performance of the models changes over time. In Figures 3 and 4, the hourly price changes indicate that there are certain periods with a lot of volatility. It is expected, that the models perform worse, especially around the times, when the market volume is increased. In Figure 5 three months with extraordinarily high spikes are displayed[11]. These

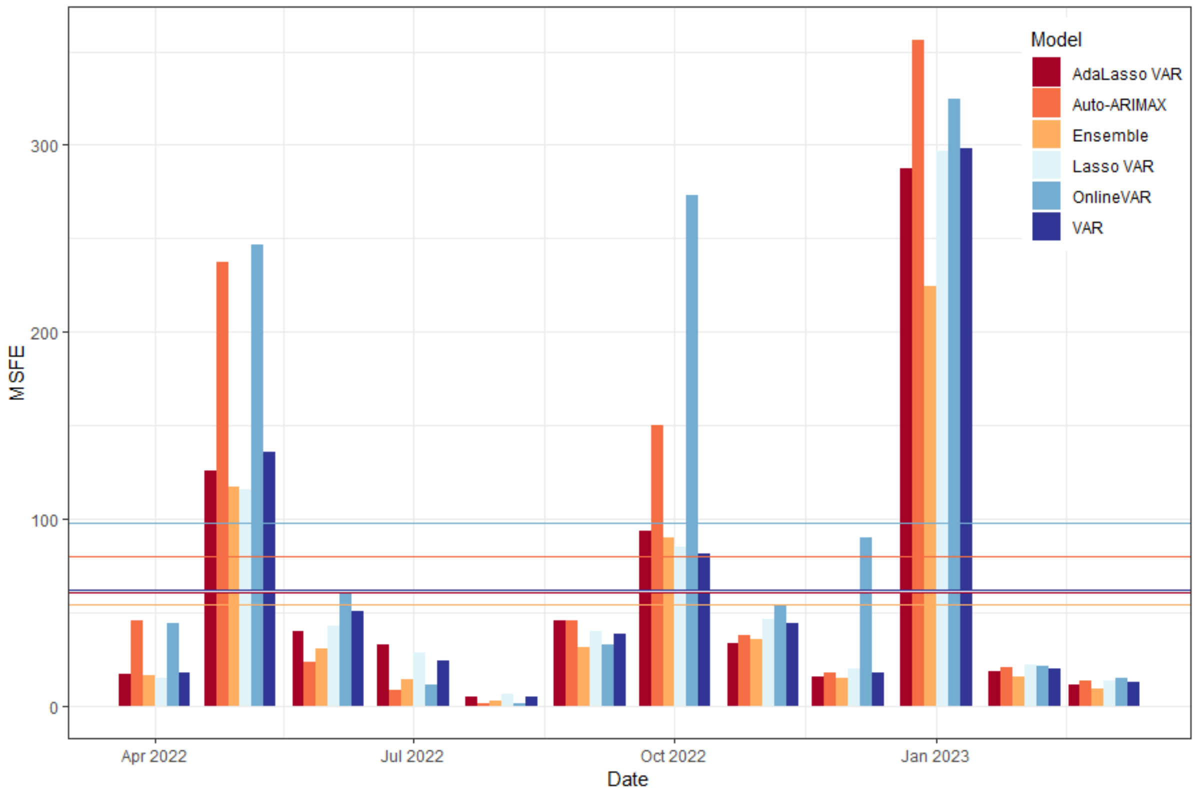

Figure 5: The performance according to the MSFE of all five models for each month in the FCR-D Down A1. The horizontal lines in the respective color represent the mean MSFE for each model reported in Tables 3 and 4.

spikes occur all around the market volume increase. The MSFE is getting particularly bad during January 2023, when the market volume increased for the last time. The mean line of the corresponding models is only exceeded in those three months. Excluding the three worst-performing months, the MSFE of the Ensemble model drops to 18.75. While this is a good indicator for a solid prediction, the volatile markets with high prices are interesting in terms of profit. To mitigate this, the OnlineVARs exponential factor is set up in a way, that the model considers the last 24h before the delivery day. This may lead to predictions with more spikes, as the model coefficients get quickly bigger if the last day was volatile. This behavior may help to hit important spikes. On the downside, the model may also predict negative prices more often. Since these are meaningless predictions, negative forecasts are set to zero, before the outlier filtering mentioned in 4.4 is applied. Most models are unaffected by the order of the outlier filtering since not all models have negative price forecasts. The riskier OnlineVAR, however, is penalized stronger since negative values are in that case replaced by zero instead of the last seen value. That is applied in all models for practical reasons, to penalize models that tend to predict lower because lower forecasts usually imply less profit. This can be seen in Figure 6, which shows the predictions of the OnlineVAR in red against the original FCR-D Down

[11] Figures 14, 15 and 16 in the Appendix show the MSFE for the other markets.

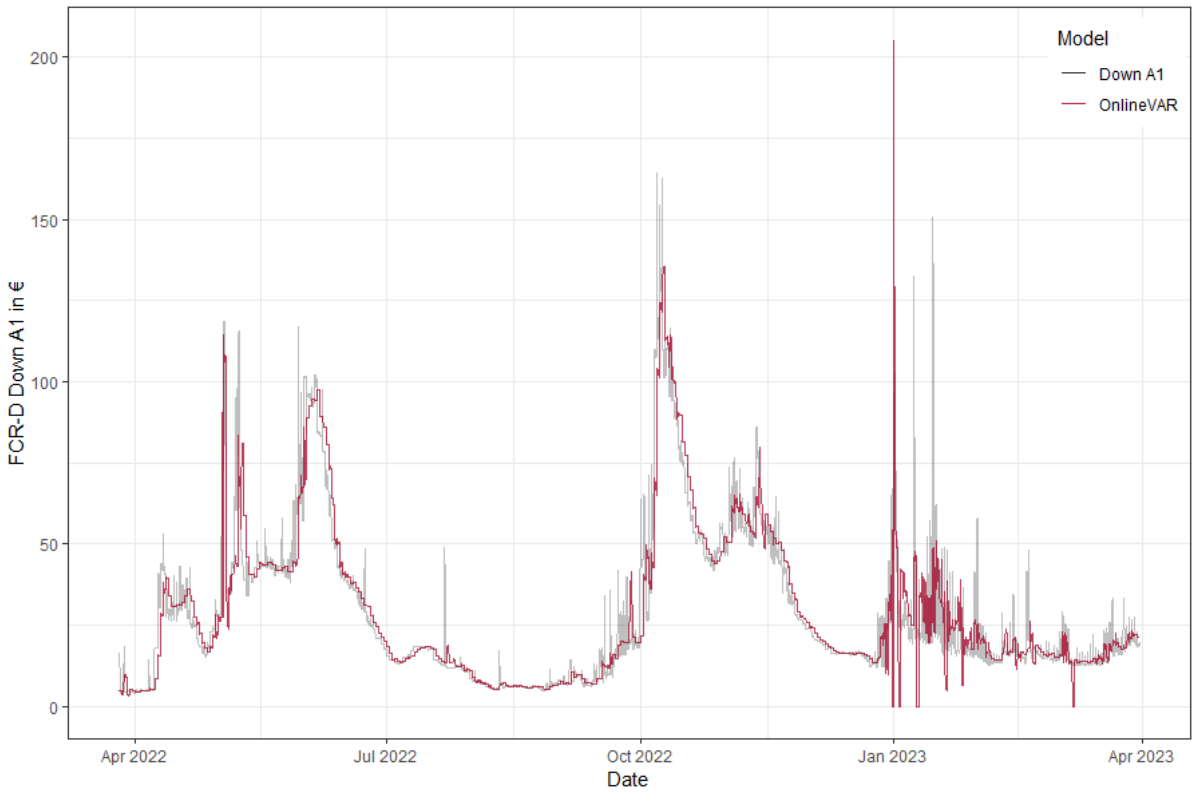

Figure 6: The predictions of the OnlineVAR compared to the original prices. Negative forecast prices are capped at zero.

A1 prices in black. The red line lags only a bit behind the original, matching most of the spikes. There are a few individual predictions, that worsen the overall performance of this model immensely. That, however, may imply, that this model alone is not reliable. On the other hand, OnlineVAR seems to be a valuable ensemble member. Including OnlineVAR reduced the overall performance of most of the ensemble predictions for the filtered and unfiltered data set.

In a similar fashion, the following two Figures 7 and 8 show the single best-performing forecast for both FCR-D Down markets, compared to the respective ensemble predictions and the original prices.

The best-performing models are chosen from Table 3 for the unfiltered data set, according to the MSFE value. For the A1 auction, this is the Adaptive Lasso model, which has the next lowest MSFE compared to the ensemble prediction. It may be challenging to identify the blue line since the red line is comparable and overlapping. When the blue line, representing the Adaptive Lasso, shines through, it is mostly above the red line, which indicates its ability to hit potential spikes on the upside. In periods with low volatility, the black line seems to be entirely covered by both forecasts. It should be noted that the root of the MSFE would give prices in Euro, meaning that in the months without May, October, and January, the prices are on average less than five Euros off.

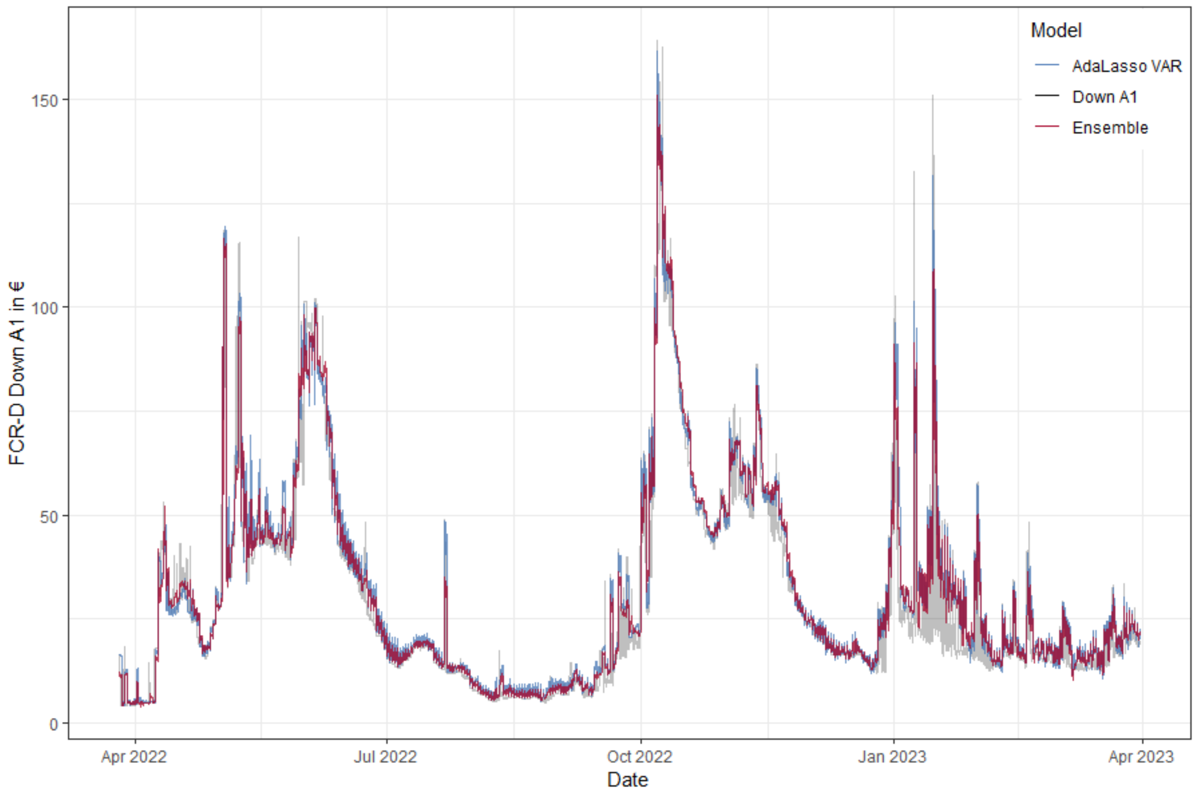

Figure 7: Comparison of the best predictors for FCR-D Down A1 prices. Negative forecast prices are capped at zero.

Similar conclusions about the performance from the ensemble (red) and the Lasso VAR model (blue) can be deducted about the Down A2 market[12]. The difference here is the price range, which is wider, and the spikes can increase up to 30 times compared to the mean value. Hitting these spikes is unrealistic. Yet, some models seem to detect the underlying pattern and can give a better indication compared to the naive model. The single best model is the Lasso VAR. It is possible, that this model is able to detect the relevant variables for predicting the spikes. Again, similar to Figure 7, the single best model is mostly above the line for the ensemble prediction, indicating that models being able to hit the spikes do perform better in terms of the chosen KPI.

Applying the Lasso VAR models allow for drawing some conclusions about variable importance. While inference on the estimated parameters of a Lasso model is usually invalid (Adamek et al., 2022), it is still possible to deduct some information about the importance of the variables. Since this is not the main focus of this work, only a brief variable importance section is presented here. It is based on the frequency of selecting the variables on different forecasting horizons, which can help to gain a better understanding of important drivers of the forecasts. Given its favorable theoretical properties, as addressed in Chapter 3.5, the AdaLasso procedure with five lags on unfiltered data

[12]Figures 17 and 18 in the Appendix contain the plots for the other predictions of the other markets.

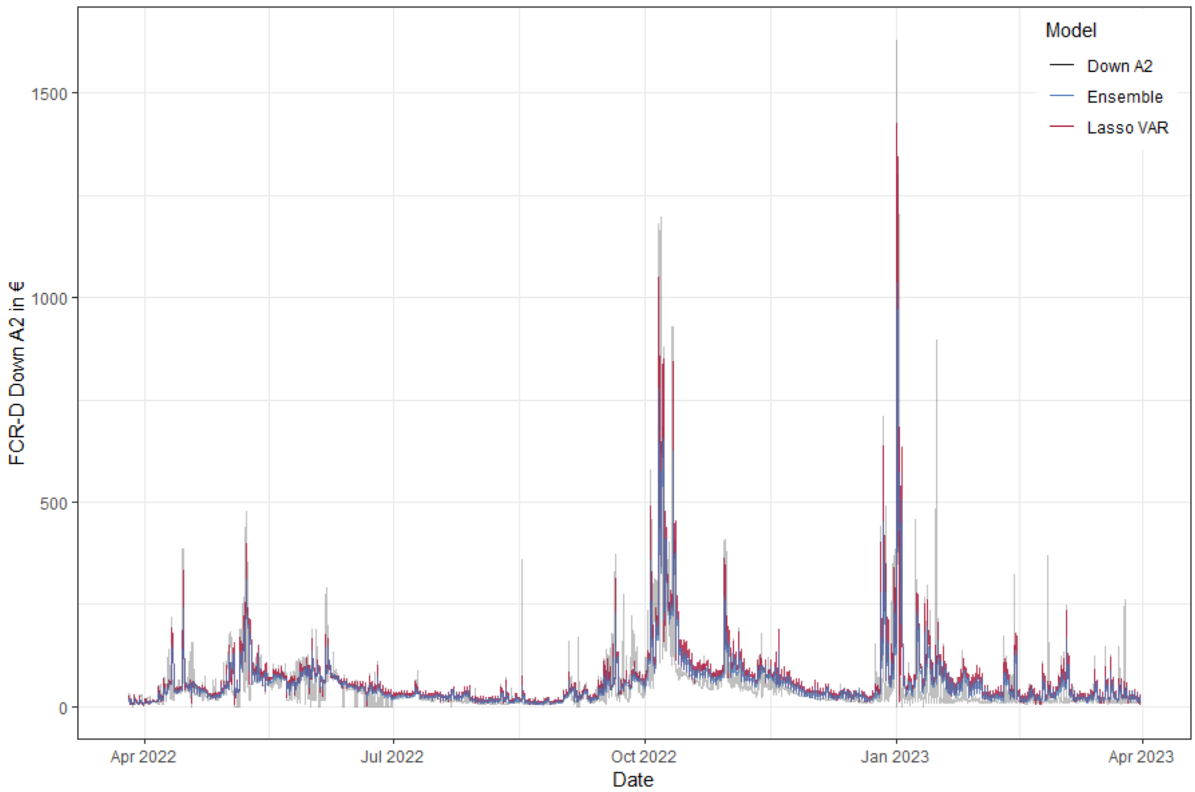

Figure 8: Comparison of the best predictors for FCR-D Down A2 prices. Negative forecast prices are capped at zero.

is used to conduct this analysis. The five most frequently selected variables of the algorithm are identified on three different forecasting horizons $h = 1, 12, 24$, with $h = 1$ being the shortest and $h = 24$ being the longest forecast horizon that has been used in the previous analysis. Since direct forecasts are used for each of the horizons, 370 iterations are conducted on each horizon. Hence, a variable that occurs 370 times has been selected throughout the full backtest period. The results for the FCR-D Down market are presented in Table 5, while the results for the FCR-D Up market can be found in Appendix 9.

The first thing to highlight is that for both FCR-D Down A1 and FCR-D Up A1 price forecasts one hour ahead only the lagged target variable is selected. This result is intuitive since for that short horizon the previous value of the target value should be very important. While for the FCR-D Down A1 price forecasts with $h = 12$ the FCR-D market data, especially the prices, seem to be of high importance, for $h = 24$ the additional variables residual load, wind forecast, and temperature seem to gain importance. For these horizons, 6.61 and 7.22 variables have been selected on average and all variables have been selected at least once for one of the horizons with one of its lags. On the more difficult-to-predict A2 market, generally, more variables have been selected on each forecast horizon ($h = 1$: 10.98, $h = 12$: 18.58, $h = 24$: 15.91). It should be noted that

Table 5: Top 5 selected variables for different forecast horizons (h) of the AdaLasso VAR model. n indicates how often a variable was selected among the total of 370 model runs. The lag of each variable is indicated in parentheses. * refers to the case when only one variable has been used in all of the models.

FCR-D Down				
	A1		**A2**	
h	**Variable**	n	**Variable**	n
1	FCR-D Down Price A1 (1L)*	370	FCR-D Down Price A2 (1L)	370
			FCR-D Down Price A2 (3L)	274
			FCR-D Down Price A2 (4L)	273
			aFRR Down Volume (1L)	253
			FCR-D Down Price A1 (1L)	248
12	FCR-D Down Price A1 (1L)	370	FCR-D Down Price A2 (1L)	370
	FCR-D Down Volume A1 (5L)	181	FCR-D Down Price A2 (5L)	345
	FCR-D Up Price A1 (1L)	156	Wind Forecast (1L)	314
	FCR-D Down Price A1 (5L)	155	FCR-D Down Price A2 (5L)	313
	aFRR Up Volume (5L)	140	Nuclear Power Production (1L)	289
24	FCR-D Down Price A1 (1L)	370	Temperature (5L)	338
	FCR-D Up Price A1 (1L)	357	FCR-D Down Price A2 (1L)	337
	Residual Load (1L)	171	FCR-D Up Price A1 (1L)	335
	Wind Forecast (5L)	170	FCR-D Down Price A1 (1L)	334
	Temperature (5L)	157	FCR-D Down Volume A2 (1L)	245

because of the lag length of five a total of 60 variables are available for the A1 market, and 85 variables on the A2 market are available. For the short and medium forecast horizons, the lagged target variable is again the most important, whereas, for the longer horizon, the temperature seems to be slightly more important. For $h = 1$ and $h = 24$, FCR-D market data is dominating the top selection, whereas for $h = 12$, also the wind forecast and nuclear power production are of importance. Often, lag size 5 of the variables is selected, which could be an indication that some values from the past are relevant for predicting future FCR-D Down prices. That could imply that the selected lag size of the model could be increased. However, given computational restrictions, that part is left for future research, and the lag size is kept as selected by the BIC.

For the A1 market, the FCR-D Up models seem to be more sparse, with fewer selected variables on average. While for $h = 12$, only 2.95 variables are selected for the FCR-D Up model (6.61 for FCR-D Down), for $h = 24$ only 1.44 (7.22) variables are selected on average. However, for the A2 market, the FCR-D models are usually sparser while selecting many more variables compared to A1. For FCR-Up 21.06 variables are selected on average for $h = 1$, 16.79 for $h = 12$, and 15.53 for $h = 24$. Analyzing the variable selection for the FCR-D Up market confirms the previous observation of the FCR-D market data being the most relevant variables for the AdaLasso forecasting model, especially on the short forecasting horizon. Other variables of importance are the temperature, the residual load, wind forecasts, and aFRR market data. While all variables have been selected at least in one model, the spot price data did not make it to the top 5 selection. It should be noted, that the results for the A2 markets should be taken with particular care as the market is difficult to predict, and the model might not be able to capture the markets' behaviors. Still, the variables outlined here can serve as an indication of which variables are of importance.

6 Research Outlook

Before concluding the analysis, some aspects relevant to potential future research will be mentioned. Following the structure of this work, the parameter choice for the filters could be optimized by performing a broader grid search, and the filters could be tuned more intensively. The SES approach, for example, could be equipped with a seasonal component (Hyndman and Athanasopoulos, 2021). Further research on filtering is required to see if a lack of tuning or information loss through filtering resulted in the worsened performance of the filtered forecasts. Another issue could be related to the variable selection. While a total of 17 variables have been considered in this work, some research in the field used many more variables (see Chapter 2). While the presented evidence suggests that the majority of the selected variables play a significant role in the forecasting process, further explanatory variables might have increased the predictability of the FCR-D prices. Data on hydropower generation could be an interesting but hard-to-obtain source for further improvements. Further meteorological data could be of help as well. However, including more data will increase the computational costs of the approaches exponentially. A solution to this could be employing the computationally efficient OnlineVAR for further high-dimensional research.

During the model selection, a forward-looking bias has been introduced by running the lag selection through the BIC on the full data set. While running the test on each iteration would have been too costly from a computational perspective, the selection used in this work seemed to be relatively robust to choosing smaller sample sizes, which is why that bias should not be considered to be a major issue. The selection of the forecasting models itself poses another challenge to tackle in future research. While this work focused on Lasso VAR applications, considering further modifications such as the Post Lasso or the general Elastic Net regularization might help to obtain better results. Using a completely different model choice, such as typical machine learning models (Decision Trees, NNs), might improve the performance further. However, it is the belief of the authors that more variables and a longer timeframe will be required for successfully applying such models. Reference models, as used in the research introduced in Chapter 2, have been tested on the data set and were found to perform worse compared to the time-series approaches. For the implementation of the forecasting algorithm, no rolling window approach was considered to limit the model comparison. However, using that method could bear a lot of potential, especially given the changes in the market structure. Those could make past observations become less important and could be captured using a properly designed rolling window approach. That involves finding the optimal window size. For the variable importance, changes over time could be interesting to research further. However, given the short past of the FCR-D market in Sweden, that might be more interesting when the market has further developed.

With regard to reporting the results, no confidence bands have been added. Adding a measure of model confidence could be helpful for further research and help for placing the bids. That topic has briefly been covered by including the ensemble model, which will result in a more confident prediction. However, that does not include any information about the possible variation of the results. As Kraft et al. (2020) highlight, the ongoing changes in the FCR-D market should always be considered when dealing with forecasts from econometric models as they are trained on past data. Structural, and technological changes, as well as behavioral changes by market participants, pose a challenge to the predictability of the market, especially given its short past. While this work has been developed, the market volume has been increased again followed by a volatile period. That means that the validity of the models needs to be assessed before applying them to the changed market.

7 Conclusion

Throughout the work, it becomes apparent that filtering the data prior to forecasting has not significantly increased the performance of the models used. While the filtering sometimes seemed to improve individual predictions of the time series, the overall performance of the models became worse. An indication may be that some relevant signals may be lost when attempting to reduce the noise. More research about noise filtering and inference is recommended to find methods that actually improve predictability. As outlined in the introduction, the expectation for better performance through more complex models is partly met. The more volatile A2 market seems to favor the simplest models, where either the naive forecast or the univariate Auto-ARIMAX is the single best model. The A1 market, on the other hand, favors penalized multivariate models, where Lasso and AdaLasso VAR are the single best models in terms of MSFE. Notably, the Auto-ARIMAX is the single best for four out of eight combinations in terms of MAE. While the MAE is a good KPI to check the robustness of the overall timeline, it is not a good indicator for models predicting spikes successfully. In addition, among the multivariate models, it is found that the VAR often yields comparable results to the more complex penalized VAR approaches. That could indicate that penalizing the models could come at the cost of removing signals in the data that can indeed be helpful for predicting FCR-D prices.

Most of the introduced models are capable of beating the naive predictor, especially for the A1 market. Following the idea of getting a more confident forecast when combining multiple models, an ensemble predictor is considered. If all models, except for the naive, are combined, there might be potential to have a model which contains the best of both worlds. Having models that are good at predicting spikes combined with models that are rather stable over time, can result in an overall better performance. The ensemble prediction is on average the best model. Combining all models significantly drops the MSFE for the Down market. For practical implications, it should be noted that using the ensemble predictions requires applying all models in parallel which will be more computationally expensive. While ensemble prediction methods have demonstrated effectiveness to enhance forecasting accuracy, there is ongoing room for improvement. Exploring alternative models or developing new approaches can lead to advancements in ensemble forecasting, potentially yielding better predictions and more reliable results.

Accurately predicting spikes in energy demand or price fluctuations is particularly important for sustainable energy providers. These companies possess the flexibility to rapidly adjust their production levels, allowing them to capitalize on market opportunities. By effectively predicting and responding to spikes, sustainable energy providers can optimize their operations, perform efficiently, and contribute to the growth and adoption of renewable energy sources. This work focused on providing forecasts of mean prices and, therefore, might not be applicable to maximizing profits in the FCR-D market. The research conducted here provides useful tools and insights that help companies make smarter decisions and optimize their operations in the market. While FCR-D Down is especially interesting to companies focusing on renewable energy, the thesis may contribute to the growth and success of sustainable energy initiatives, promoting a greener and more sustainable future.

References

Adamek, R., Smeekes, S., & Wilms, I. (2022). Lasso inference for high-dimensional time series. *Journal of Econometrics*. https://doi.org/https://doi.org/10.1016/j.jeconom.2022.08.008

Afanasyev, D. O., & Fedorova, E. A. (2019). On the impact of outlier filtering on the electricity price forecasting accuracy. *Applied Energy*, *236*, 196–210. https://doi.org/https://doi.org/10.1016/j.apenergy.2018.11.076

Auray, S., & Caponi, V. (2020). *A Vector Autoregressive Model of Forecast Electricity Consumption in France* (Working Papers No. 2020-06). Center for Research in Economics and Statistics. https://ideas.repec.org/p/crs/wpaper/2020-06.html

Badhwar, G., & Sappal, A. S. (2016). Noise reduction using kalman filter. *International journal of Latest Trends in Engineering &amp Technology*, *7*(1). https://doi.org/10.21172/1.71.091

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation [Data Mining for Software Trustworthiness]. *Information Sciences*, *191*, 192–213. https://doi.org/https://doi.org/10.1016/j.ins.2011.12.028

Bühlmann, P., & van de Geer, S. (2011). *Statistics for high-dimensional data*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20192-9

Callot, L. A., & Kock, A. B. (2014). Oracle efficient estimation and forecasting with the adaptive lasso and the adaptive group lasso in vector autoregressions. In N. Haldrup, M. Meitz, & P. Saikkonen (Eds.), *Essays in nonlinear time series econometrics* (pp. 238–266). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679959.003.0010

Callot, L. A., Kock, A. B., & Medeiros, M. C. (2014). *Estimation and forecasting of large realized covariance matrices and portfolio choice*. https://doi.org/10.2139/ssrn.2523848

Cavalcante, L., Bessa, R. J., Reis, M., & Browell, J. (2016). LASSO vector autoregression structures for very short-term wind power forecasting. *Wind Energy*, *20*(4), 657–675. https://doi.org/10.1002/we.2029

Cavalcante, L., Bessa, R. J., Reis, M., & Browell, J. (2017). Lasso vector autoregression structures for very short-term wind power forecasting. *Wind Energy*, *20*(4), 657–675. https://doi.org/https://doi.org/10.1002/we.2029

Choi, I., & Jeong, H. (2020). Differencing versus nondifferencing in factor-based forecasting. *Journal of Applied Econometrics*, *35*(6), 728–750. https://doi.org/10.1002/jae.2777

Davis, R. A., Zang, P., & Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, *25*(4), 1077–1096. https://doi.org/10.1080/10618600.2015.1092978

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–431. Retrieved May 17, 2023, from http://www.jstor.org/stable/2286348

Dowell, J., & Pinson, P. (2016). Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Transactions on Smart Grid*, *7*(2), 763–770. https://doi.org/10.1109/TSG.2015.2424078

Enders, W. (2015). *Applied econometric time series* (Fourth edition). Wiley.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22.

García-Ascanio, C., & Maté, C. (2010). Electric power demand forecasting using interval time series: A comparison between var and imlp. *Energy Policy*, *38*(2), 715–725. https://doi.org/https://doi.org/10.1016/j.enpol.2009.10.007

Gilmore, J., Nolan, T., & Simshauser, P. (2022). *The Levelised Cost of Frequency Control Ancillary Services in Australiaâ€™s National Electricity Market* (Cambridge Working Papers in Economics No. 2203). Faculty of Economics, University of Cambridge. https://ideas.repec.org/p/cam/camdae/2203.html

Giovanelli, C., Sierla, S., Ichise, R., & Vyatkin, V. (2018). Exploiting artificial neural networks for the prediction of ancillary energy market prices. *Energies*, *11*(7). https://doi.org/10.3390/en11071906

Haldrup, N., Nielsen, F. S., & Nielsen, M. Ø. (2010). A vector autoregressive model for electricity prices subject to long memory and regime switching. *Energy Economics*, *32*(5), 1044–1058. https://doi.org/https://doi.org/10.1016/j.eneco.2010.02.012

Hameed, Z., Pollitt, M., Kattuman, P., & TrÃ¦holt, C. (2023). *Frequency markets and the problem of pre-dictability* (Cambridge Working Papers in Economics No. 2306). Faculty of Economics, University of Cambridge. https://ideas.repec.org/p/cam/camdae/2306.html

Hanzák, T., Tomáš; Cipra. (2011). Exponential smoothing for time series with outliers. *Kybernetika*, *47*(2), 165–178. http://eudml.org/doc/197187

Hao, K., Chen, Z., Yang, Q., & Xie, S. (2022). A fast and efficient short-term wind speed forecast model for wind turbines. *2022 IEEE 10th International Conference on Smart City and Informatization (iSCI)*, 35–40. https://doi.org/10.1109/iSCI57775.2022.00015

Haykin, S. (2002). *Adaptive filter theory* (4th). Prentice Hall.

He, M., Vittal, V., & Zhang, J. (2015). A sparsified vector autoregressive model for short-term wind farm power forecasting. *2015 IEEE Power Energy Society General Meeting*, 1–5. https://doi.org/10.1109/PESGM.2015.7285972

Hecq, A., Margaritella, L., & Smeekes, S. (2021). Granger Causality Testing in High-Dimensional VARs: A Post-Double-Selection Procedure* [nbab023]. *Journal of Financial Econometrics*. https://doi.org/10.1093/jjfinec/nbab023

Hsu, N.-J., Hung, H.-L., & Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics Data Analysis*, *52*(7), 3645–3657. https://doi.org/https://doi.org/10.1016/j.csda.2007.12.004

Hurta, A., Žilka, M., & Freiberg, F. (2022). Impact of the splitting of the german–austrian electricity bidding zone on investment in a grid-scale battery energy storage system deployed for price arbitrage with gray and green power in austrian and german day-ahead power markets. *Energy Reports*, *8*, 12045–12062. https://doi.org/https://doi.org/10.1016/j.egyr.2022.09.045

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd edition). OTexts.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, *27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

Janczura, J., Trück, S., Weron, R., & Wolff, R. C. (2013). Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics*, *38*, 96–110. https://doi.org/https://doi.org/10.1016/j.eneco.2013.03.013

Kalman, R. E., et al. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*(1), 35–45.

Kamen, E. W., & Su, J. K. (1999). *Introduction to optimal estimation*. Springer London. https://doi.org/10.1007/978-1-4471-0417-9

Kim, J., & Choi, I. (2017). Unit roots in economic and financial time series: A re-evaluation at the decision-based significance levels. *Econometrics*, *5*(3), 41. https://doi.org/10.3390/econometrics5030041

Kock, A. B., & Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions [High Dimensional Problems in Econometrics]. *Journal of Econometrics*, *186*(2), 325–344. https://doi.org/https://doi.org/10.1016/j.jeconom.2015.02.013

Kock, A. B., & Teräsvirta, T. (2016). Forecasting macroeconomic variables using neural network models and three automated model selection techniques. *Econometric Reviews*, *35*(8-10), 1753–1779. https://doi.org/10.1080/07474938.2015.1035163

Kraft, E., Keles, D., & Fichtner, W. (2020). Modeling of frequency containment reserve prices with econometrics and artificial intelligence. *Journal of Forecasting*, *39*(8), 1179–1197. https://doi.org/10.1002/for.2693

Kraftnät, S. (2023a). Information on different ancillary services. https://www.svk.se/en/stakeholders-portal/electricity-market/provision-of-ancillary-services/information-on-different-ancillary-services/

Kraftnät, S. (2023b). Information on different ancillary services. https://www.svk.se/en/stakeholders-portal/electricity-market/provision-of-ancillary-services/information-on-different-ancillary-services/

Lago, J., Marcjasz, G., De Schutter, B., & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, *293*, 116983. https://doi.org/https://doi.org/10.1016/j.apenergy.2021.116983

Luetkepohl, H. (2011). *Vector autoregressive models* (Economics Working Papers ECO2011/30). European University Institute. https://ideas.repec.org/p/eui/euiwps/eco2011-30.html

Ma'arif, A., Iswanto, I., Nuryono, A. A., & Alfian, R. I. (2019). Kalman filter for noise reducer on sensor readings. *Signal and Image Processing Letters*, *1*(2), 11–22. https://doi.org/10.31763/simple.v1i2.2

Messner, J. W., & Pinson, P. (2019). Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *International Journal of Forecasting*, *35*(4), 1485–1498. https://doi.org/https://doi.org/10.1016/j.ijforecast.2018.02.001

Narajewski, M. (2022). Probabilistic forecasting of german electricity imbalance prices. *Energies*, *15*(14). https://doi.org/10.3390/en15144976

Narajewski, M., & Ziel, F. (2020). Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets*, *19*, 100107. https://doi.org/https://doi.org/10.1016/j.jcomm.2019.100107

Nicholson, W. B., Wilms, I., Bien, J., & Matteson, D. S. (2020). High dimensional forecasting via interpretable vector autoregression.

Orfanidis, S. J. (2018). *Applied optimum signal processing* (2018th ed.). McGraw-Hill.

Pfaff, B. (2008). Var, svar and svec models: Implementation within R package vars. *Journal of Statistical Software*, *27*(4). https://www.jstatsoft.org/v27/i04/

Pihl, H. (2019).

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.

Shah, I., Akbar, S., Saba, T., Ali, S., & Rehman, A. (2021). Short-term forecasting for the electricity spot prices with extreme values treatment. *IEEE Access*, *9*, 105451–105462. https://doi.org/10.1109/ACCESS.2021.3100076

Siddiqui, A., Marnay, C., & Khavkin, M. (2001). Spot pricing of electricity and ancillary services in a competitive california market. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, 9 pp.-. https://doi.org/10.1109/HICSS.2001.926290

Song, S., & Bickel, P. J. (2011). Large vector auto regressions.

Swanson, N. R., & White, H. (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, *13*(3), 265–275. https://doi.org/10.1080/07350015.1995.10524600

Taieb, S. B., Bontempi, G., Atiya, A., & Sorjamaa, A. (2011). A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*(4), 1030–1081. https://doi.org/https://doi.org/10.1016/j.ijforecast.2014.08.008

Zhao, P., & Yu, B. (2006). On model selection consistency of lasso. *Journal of Machine Learning Research*, *7*(90), 2541–2563. http://jmlr.org/papers/v7/zhao06a.html

Zivot, E., & Wang, J. (2003). Vector autoregressive models for multivariate time series. In *Modeling financial time series with s-plus®* (pp. 369–413). Springer New York. https://doi.org/10.1007/978-0-387-21763-5_11

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. https://doi.org/10.1198/016214506000000735

A Appendix

Auction 2 Timeline

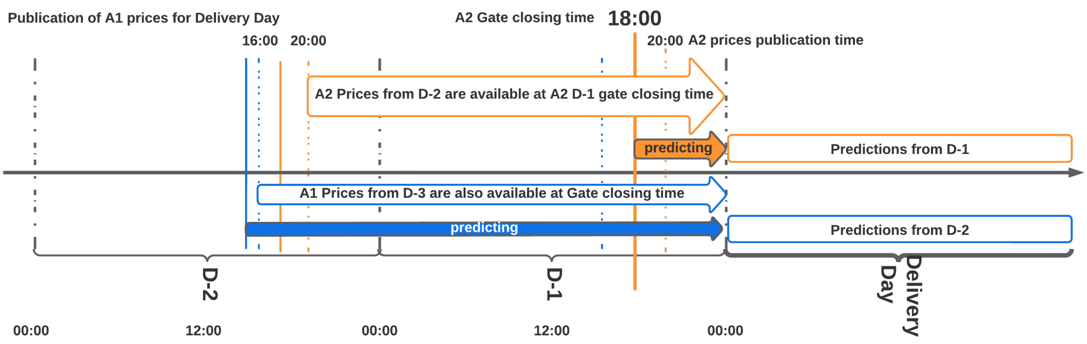

Figure 9: The timeline highlights important times for the second auction A2. The bold orange vertical line at 18:00h shows the gate closing time, while the dashed line two hours later indicates the results of the auction. The blue colors represent data or related points to A1 prices, while orange represents elements of A2.

FCR-D Up Heatmaps

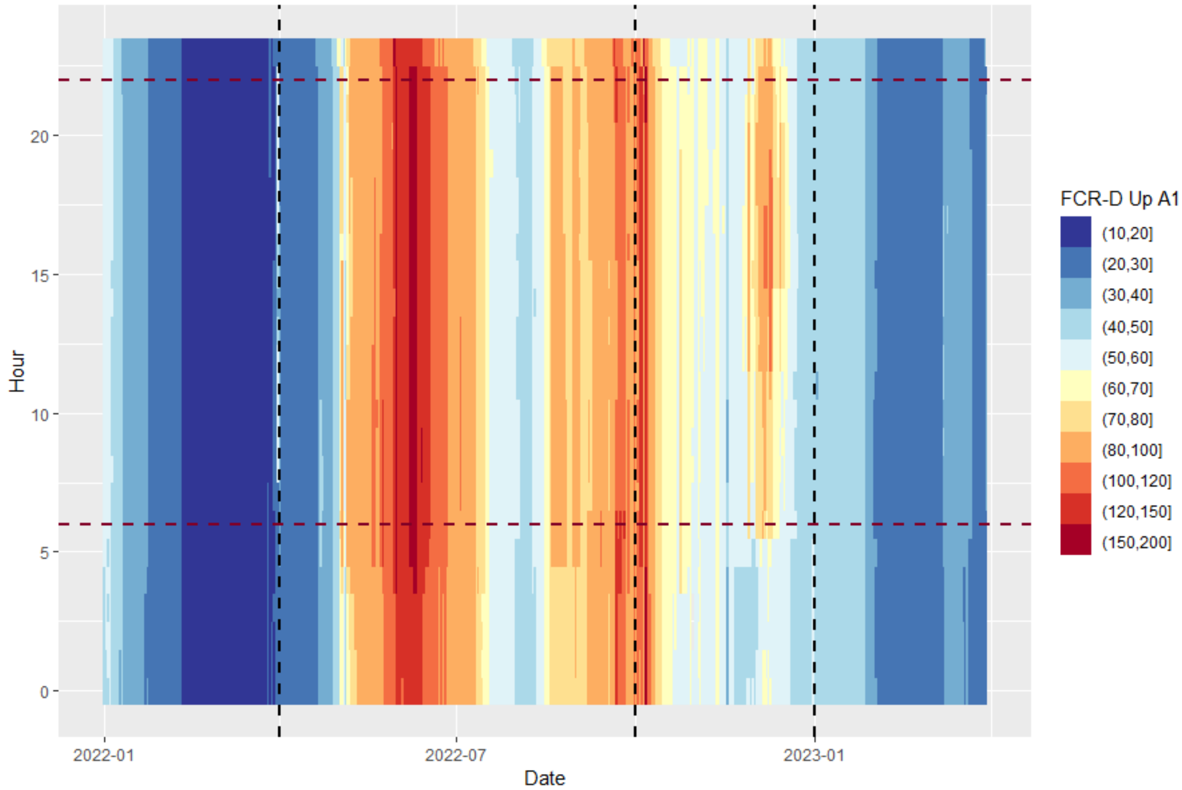

Figure 10: The heatmap shows the prices of FCR-D up (2 Day ahead auction) for each day over the entire timespan. The horizontal red lines indicate the switches between morning, day and evening, while the purple lines highlight the calendarial seasonal changes and the pink vertical lines indicate the time when the market volume of the product was increased. Note that the white spots around June 2022 are due to very strong outliers that distort the coloring of the entire heatmap.

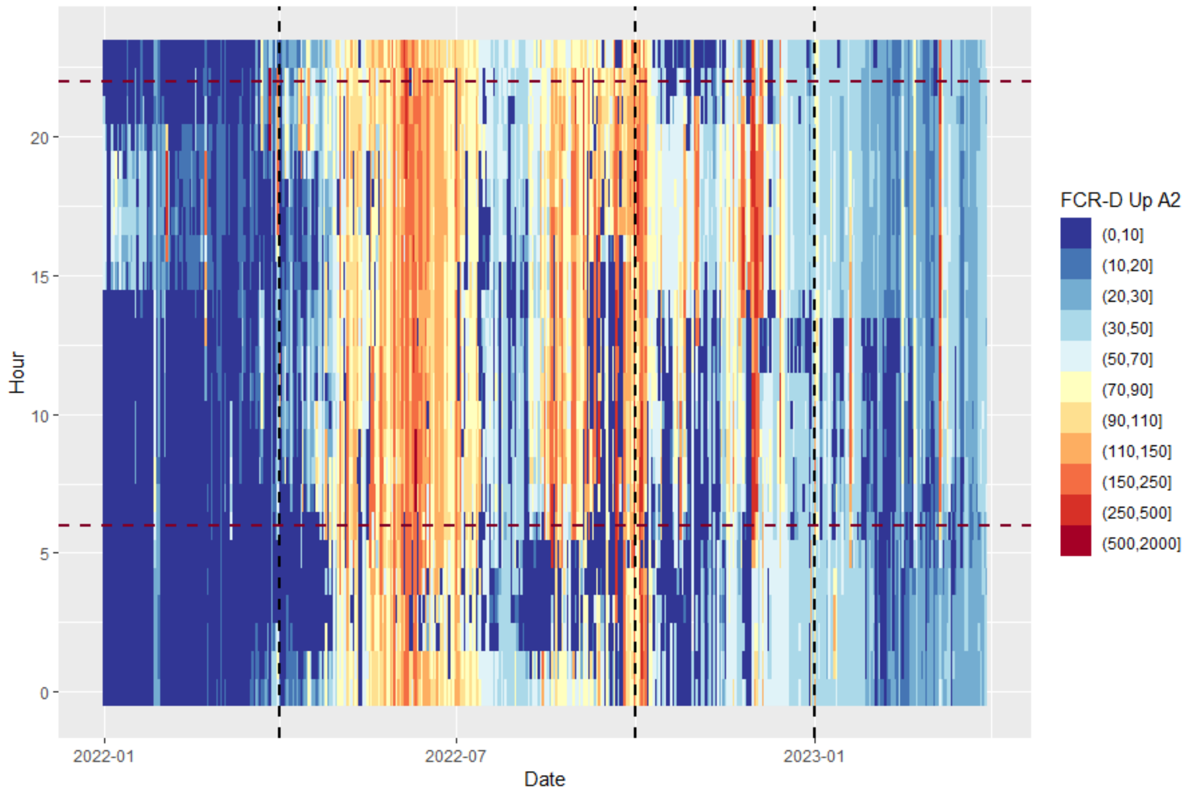

Figure 11: The heatmap shows the prices of FCR-D up (1 Day ahead auction) for each day over the entire timespan. The horizontal red lines indicate the switches between morning, day and evening, while the purple lines highlight the calendarial seasonal changes and the pink vertical lines indicate the time when the market volume of the product was increased. Note that some strong outliers that distort the coloring of the heatmap have been removed.

Descriptive Statistics

Table 6: Descriptive statistics for all variables. N = 10,893.

Statistic	Mean	St. Dev.	Min	Max
FCR-D Down Price A1	26.07	23.54	4.13	164.28
FCR-D Up Price A1	55.77	33.13	10.25	173.86
FCR-D Down Price A2	43.36	89.20	0	1,630.15
FCR-D Up Price A2	52.42	58.67	0	1,060.09
FCR-D Down Volume A1	120.78	43.90	39.10	240.60
FCR-D Up Volume A1	478.08	42.65	68.60	599.20
FCR-D Down Volume A2	35.39	23.78	0	120
FCR-D Up Volume A2	58.76	51.60	0	542
aFRR Down Price aggregated	43.58	77.74	0	4,320
aFRR Up Price aggregated	55.78	126.91	0	6,120
aFRR Down Volume aggregated	98.27	51.50	0	241
aFRR Up Volume aggregated	83.89	49.21	0	189
Wind Production Forecast	0.32	0.25	0	0.94
Nuclear Power Production (Mio.)	5.70	0.76	3.73	6.94
Temperature	19.63	7.95	1	46.50
Spot Price SE2	656.95	806.73	−22.57	6,590.85
Residual Load (k)	11.44	3.42	−0.92	22.28

Data Sources

Table 7: Variable list and their reference in the academic literature in related works. The numbers in the source column refer to the list below. aFRR Variables marked with a * have been modified. Since the aFRR data is only available per region, the total volume has been summed up and the price averaged per time point.

Variable	Reference	Source
FCR-D Market		
FCR-D Down Result Price A1	Giovanelli et al., 2018 Hameed et al., 2023 Kraft et al., 2020 Pihl, 2019	1
FCR-D Up Result Price A1	Giovanelli et al., 2018 Hameed et al., 2023 Kraft et al., 2020 Pihl, 2019	1
FCR-D Down Result Price A2	Giovanelli et al., 2018 Hameed et al., 2023 Kraft et al., 2020 Pihl, 2019	1
FCR-D Up Result Price A2	Giovanelli et al., 2018 Hameed et al., 2023 Kraft et al., 2020 Pihl, 2019	1
FCR-D Down Result Volume A1	Giovanelli et al., 2018 Pihl, 2019	1
FCR-D Up Result Volume A1	Giovanelli et al., 2018 Pihl, 2019	1
FCR-D Down Result Volume A2	Giovanelli et al., 2018 Pihl, 2019	1
FCR-D Up Result Volume A2	Giovanelli et al., 2018 Pihl, 2019	1
aFRR Market		
aFRR Down Price aggregated*	Pihl, 2019	2
aFRR Up Price aggregated*	Pihl, 2019	2
aFRR Down Volume aggregated*	Pihl, 2019	2
aFRR Up Volume aggregated*	Pihl, 2019	2
Production		

Wind Production Forecast	Giovanelli et al., 2018 Narajewski, 2022 Pihl, 2019	3
Nuclear Power Production	Giovanelli et al., 2018 Narajewski, 2022 Pihl, 2019	4
Other		
Temperature	Giovanelli et al., 2018	5
Spot Price SE2	Kraft et al., 2020 Giovanelli et al., 2018 Narajewski, 2022 Pihl, 2019	6
Residual Load	Kraft et al., 2020 Giovanelli et al., 2018 Narajewski, 2022	7

1. Svenska Kraftnät. (2023a). FCR. Sundbyberg. Retrieved March 31, 2023, from https://mimer.svk.se/PrimaryRegulation/PrimaryRegulationIndex.

2. Svenska Kraftnät. (2023b). Automatisk frekvensåterställningsreserv. Sundbyberg. Retrieved March 31, 2023, from https://mimer.svk.se/AutomaticFrequencyRestorationReserve.

3. Energy Quantified. (2023b). SE Median Wind Forecast. Oslo. Retrieved March 31, 2023 (paid database).

4. Svenska Kraftnät. (2023c). Produktionsstatistik. Sundbyberg. Retrieved April 07, 2023, from https://mimer.svk.se/ProductionConsumption/ProductionIndex

5. Sveriges Meteorologiska och Hydrologiska Institut. (2023). Stockholm-Observatoriekullen A. Norrköping. Retrieved April 07, 2023, from https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer/#param=airtemperatureInstant,stations=core,stationid=98230

6. eSett. (2023). Spot SEK SE2. Helsinki. Retrieved March 31, 2023 (paid database).

7. Energy Quantified. (2023c). SE Residual Load Actual. Oslo. Retrieved March 31, 2023 (paid database).

Correlation Matrix

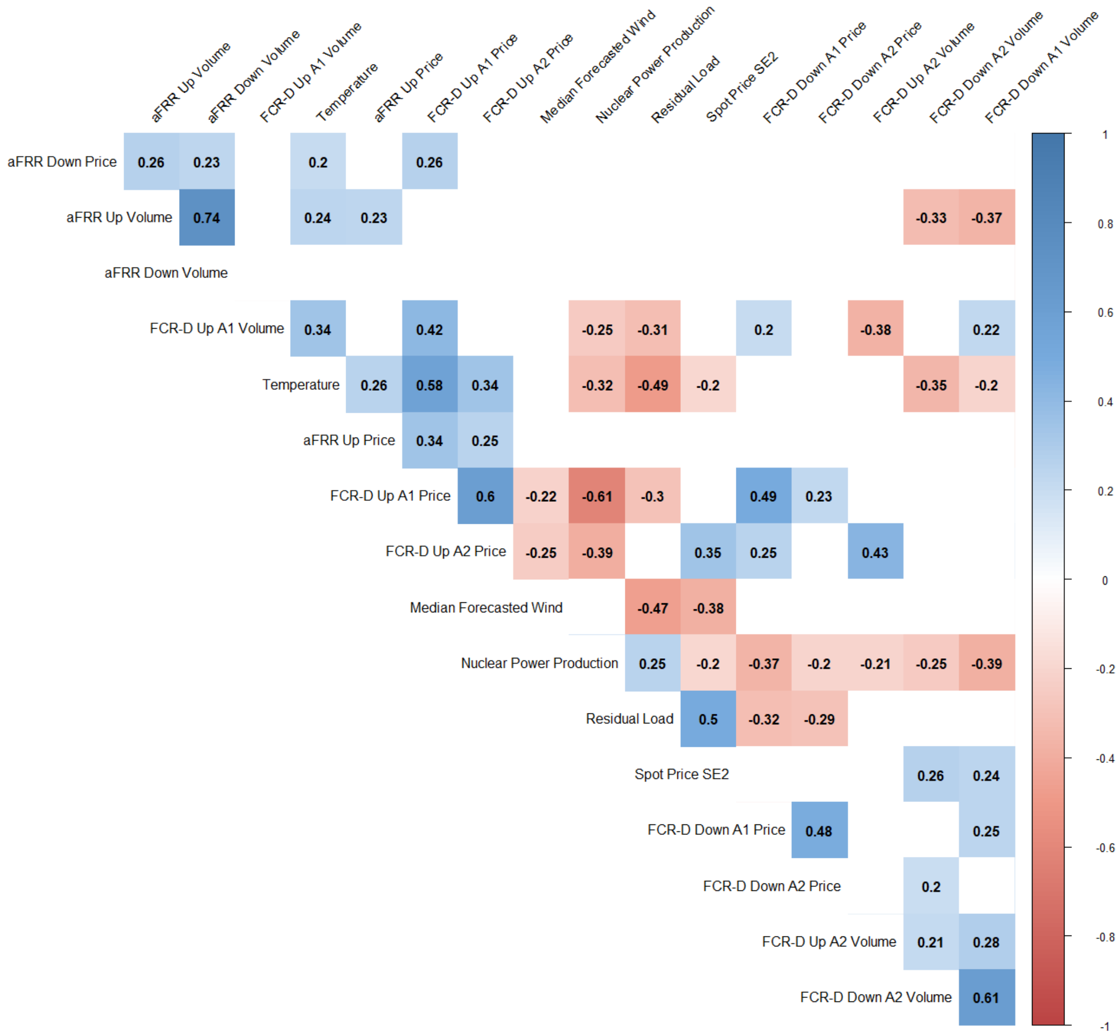

Figure 12: Correlation matrix for the included variables. For better readability, the values between −0.2 and 0.2 have been removed from the overview.

Fitting FCR-D Up Price A1

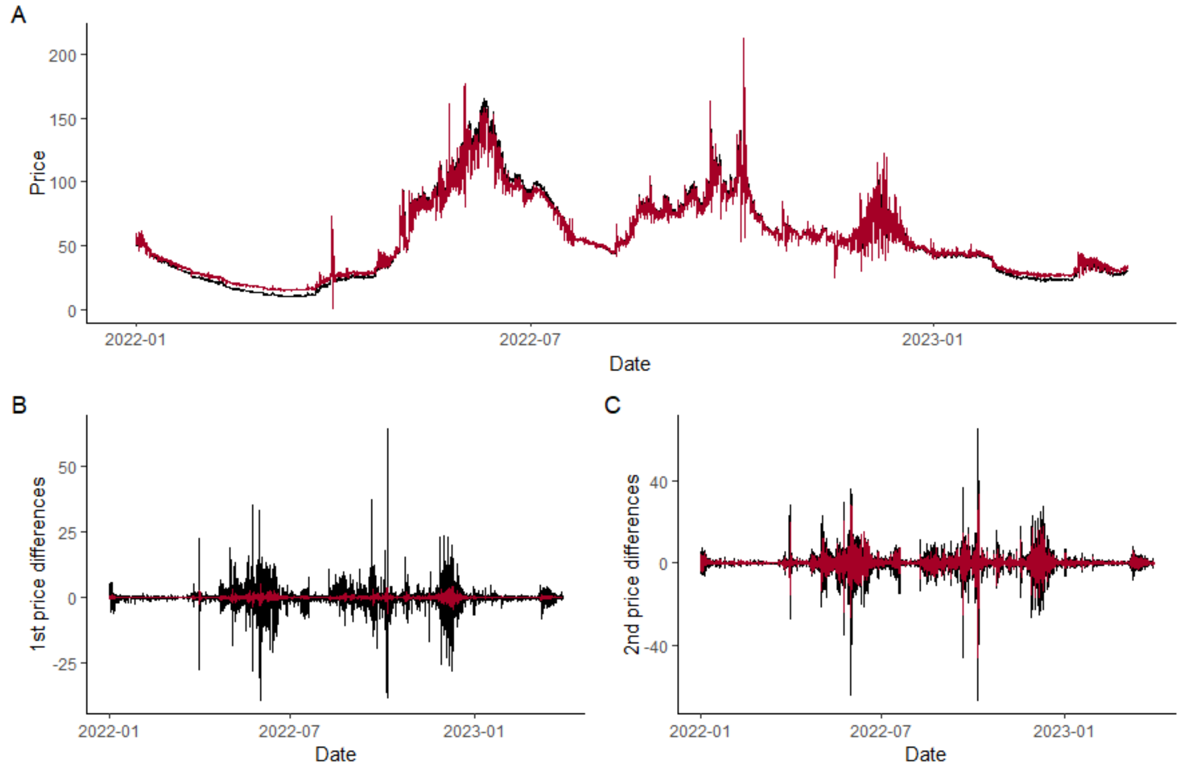

Figure 13: Comparison of three fitted auto-ARIMA models. The auto-ARIMA is constrained in a way, that it is not allowed to take differences, meaning that the differences are made prior to feeding the model. A on the top, uses the raw data and shows a relatively good fit. B on the bottom right depicts the very poor fit of first differenced time-series while the in C, the second differences show an improvement, but yet fail to fit sufficiently.

Noise Filter Performance

Table 8: Noise filter performance on all variables. For each variable, the out-of-sample performance of the Auto-ARIMA using the best-performing filter setup is compared to an Auto-ARIMA using the original series. The performance values are compared in percentage based on the MSFE (for FCR-D prices) and MAE (other variables). For the optimal filter, K refers to the Kalman Filter and F to Frequency Averaging with the respective parameter choice in parenthesis. For the Kalman Filter parameter values for R^2 between 0.1 and 1.5 have been considered, whereas for Frequency Averaging a frequency of 2 has been used.

Variable	**Performance in %**	**Optimal Filter**
FCR-D Market		
FCR-D Down Price A1	+0.85	K (0.25)
FCR-D Up Price A1	+21.05	K (0.75)
FCR-D Down Price A2	+0.66	K (0.25)
FCR-D Up Price A2	−1.56	F (2)
FCR-D Down Volume A1	+1.33	K (0.1)
FCR-D Up Volume A1	+0.97	K (0.1)
FCR-D Down Volume A2	−1.17	K (0.1)
FCR-D Up Volume A2	+0.25	K (0.1)
aFRR Market		
aFRR Down Price aggregated	+1.42	K (1.5)
aFRR Up Price aggregated	+4.55	K (0.25)
aFRR Down Volume aggregated	−2.72	K (0.5)
aFRR Up Volume aggregated	+26.58	K (0.25)
Production		
Wind Production Forecast	+0.90	K (0.1)
Nuclear Power Production	−0.84	K (0.1)
Other		
Temperature	−1.79	K (0.1)
Spot Price SE2	+1.74	K (0.1)
Residual Load	−10.89	K (0.25)

Monthly Model Performance Evaluation

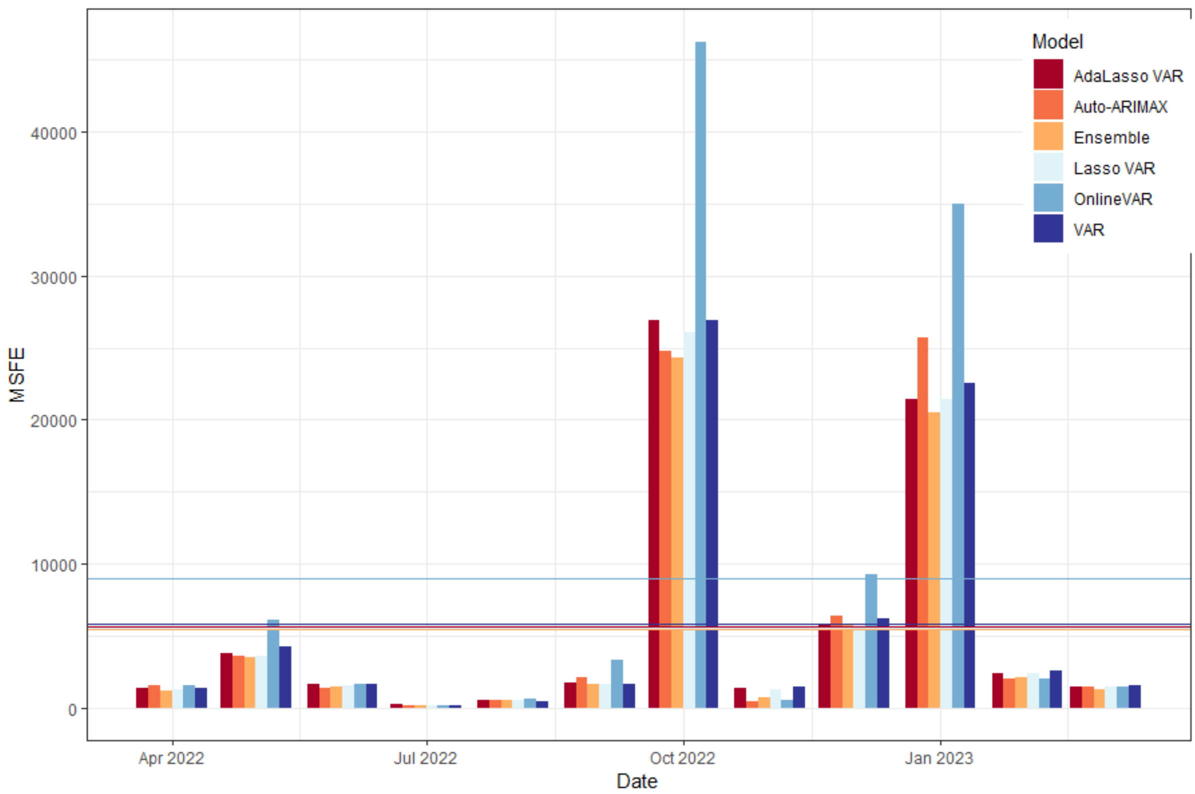

Figure 14: The performance according to the MSFE of all five models for each month in the FCR-D Down A2. The horizontal lines in the respective color represent the mean MSFE for each model reported in Tables 3 and 4.

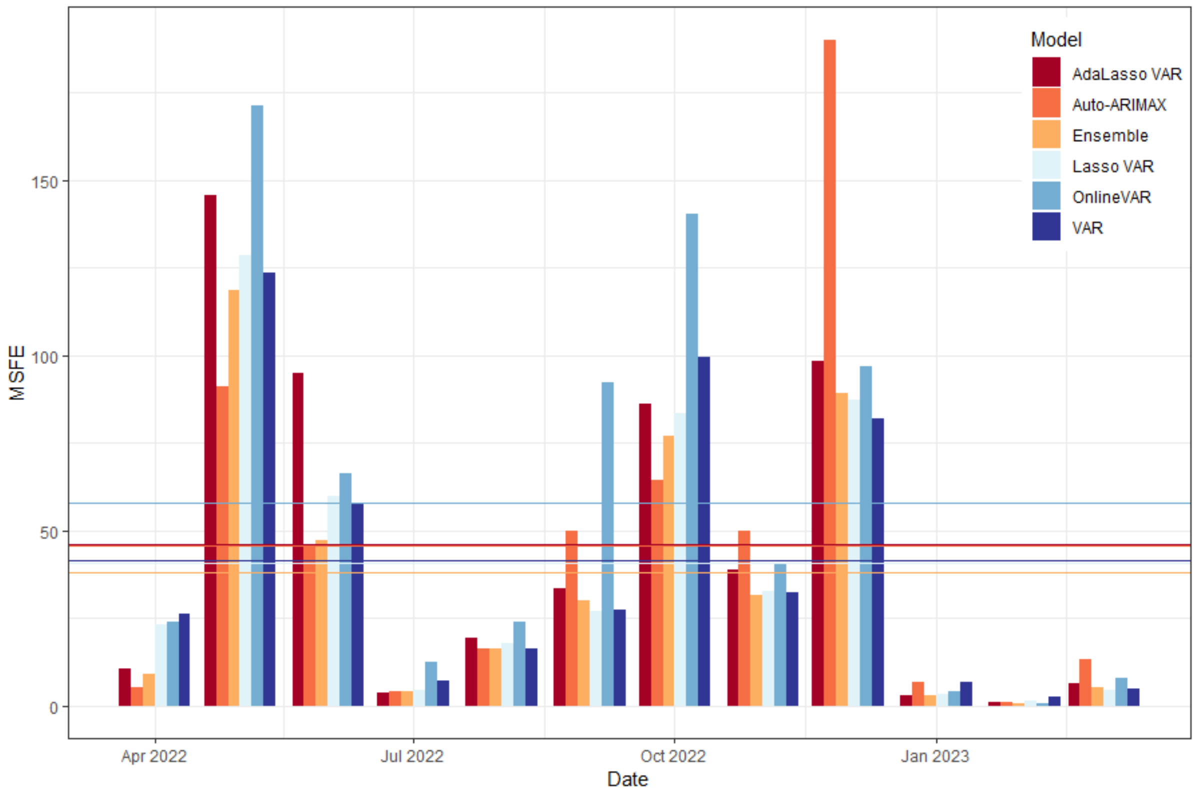

Figure 15: The performance according to the MSFE of all five models for each month in the FCR-D Up A1. The horizontal lines in the respective color represent the mean MSFE for each model reported in Tables 3 and 4.

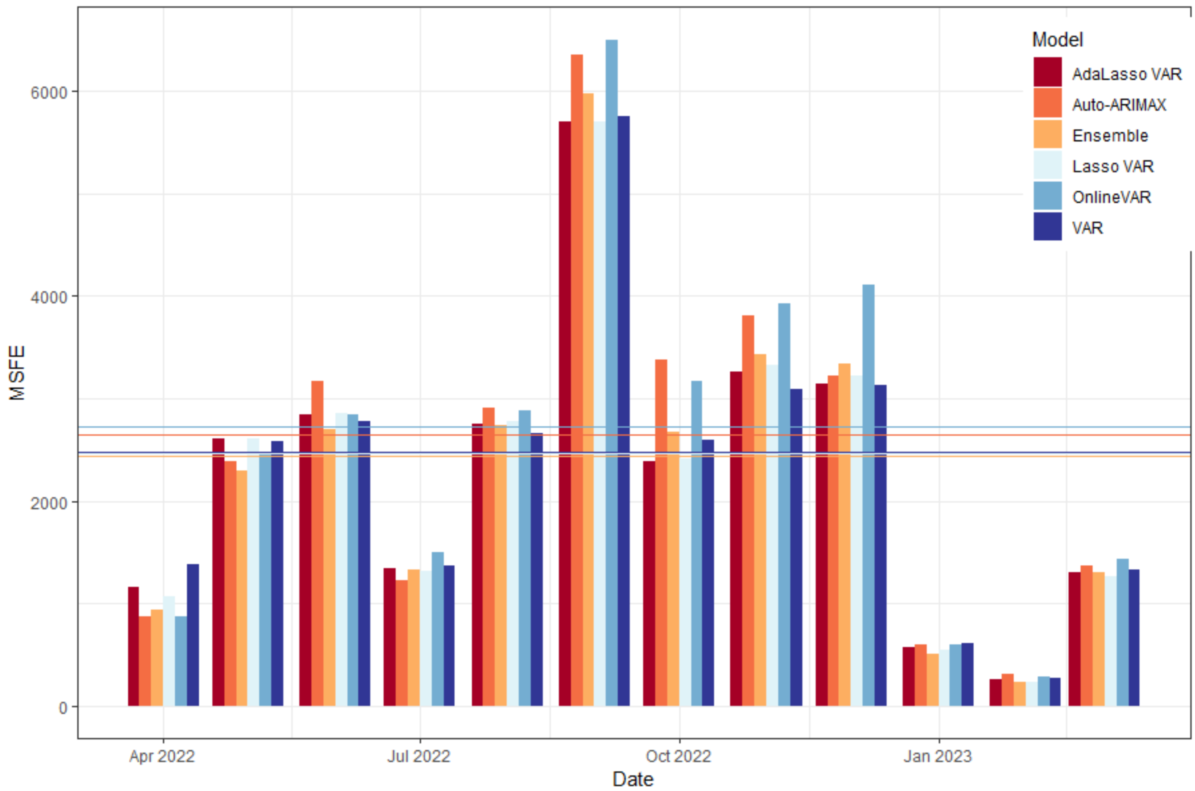

Figure 16: The performance according to the MSFE of all five models for each month in the FCR-D Up A2. The horizontal lines in the respective color represent the mean MSFE for each model reported in Tables 3 and 4.

Forecasting Plot Best Predictors

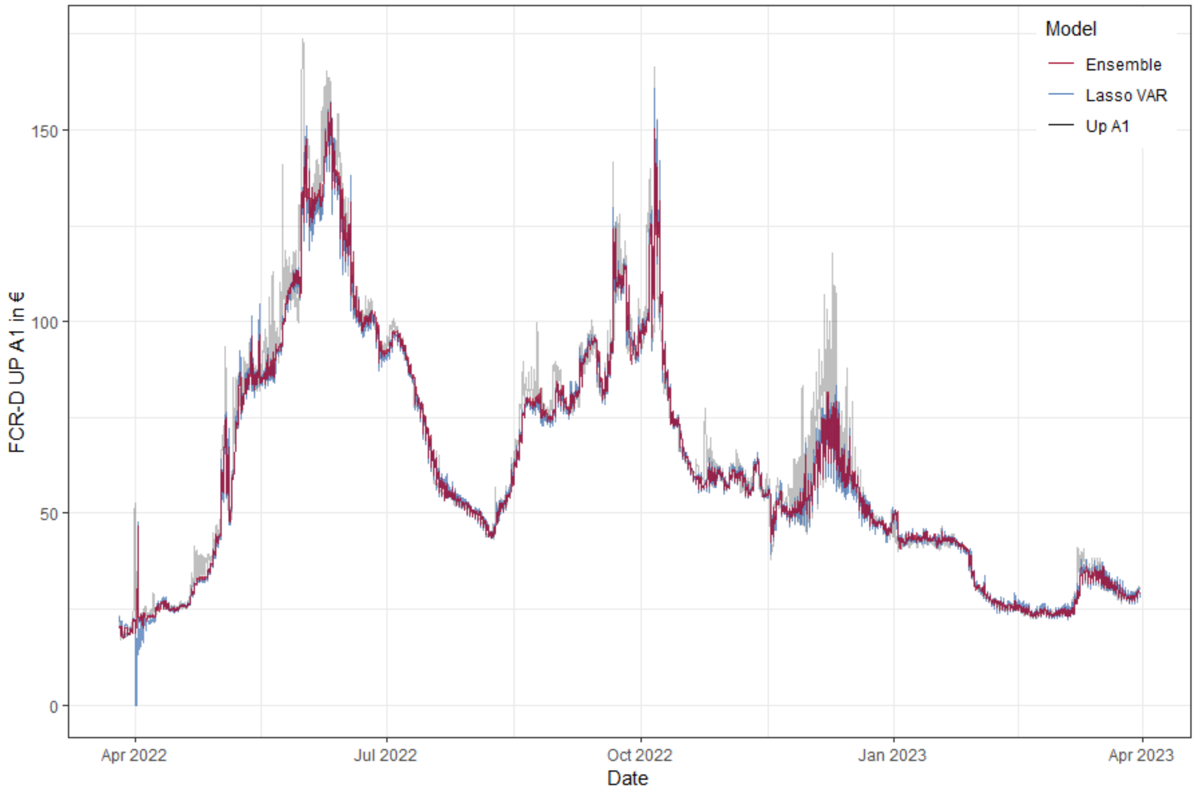

Figure 17: Comparison of the best predictors for FCR-D Up A1 prices. Negative forecast prices are capped at zero.

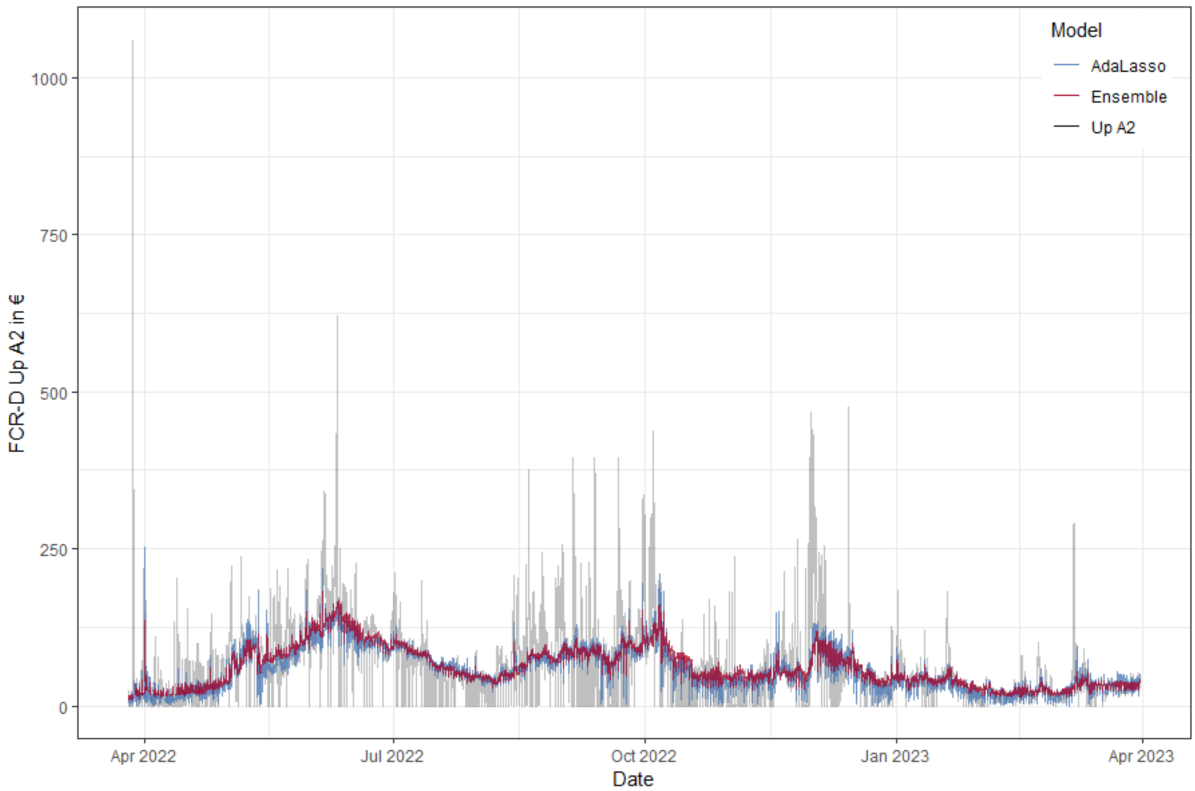

Figure 18: Comparison of the best predictors for FCR-D Up A2 prices. Negative forecast prices are capped at zero.

Variable Importance FCR-D Up Price

Table 9: Top 5 selected variables for different forecast horizons (h) of the AdaLasso VAR model. n indicates how often a variable was selected among the total of 370 model runs. The lag of each variable is indicated in parentheses. * refers to the case when only one variable has been used in all of the models.

FCR-D Up				
	A1		**A2**	
h	**Variable**	***n***	**Variable**	***n***
1	FCR-D Up Price A1 (1L)*	370	FCR-D Up Price A2 (1L)	370
			FCR-D Up Price A2 (3L)	370
			FCR-D Up Volume A2 (1L)	370
			FCR-D Up Price A2 (2L)	368
			FCR-D Up Price A2 (5L)	367
12	FCR-D Up Price A1 (1L)	370	aFRR Up Volume (5L)	369
	FCR-D Up Price A1 (5L)	218	FCR-D Up Price A1 (1L)	325
	Temperature (5L)	89	Wind Forecast (5L)	318
	Residual Load (1L)	41	FCR-D Up Price A2 (1L)	303
	Wind Forecast (1L)	34	FCR-D Up Price A2 (4L)	278
24	FCR-D Up Price A1 (1L)	370	FCR-D Up Volume A2 (1L)	368
	FCR-D Up Price A1 (5L)	21	Temperature (5L)	366
	FCR-D Down Price A1 (5L)	20	FCR-D Up Price A1 (1L)	312
	aFRR Up Price (5L)	20	FCR-D Up Price A2 (1L)	306
	Temperature (5L)	16	Wind Forecast (1L)	290