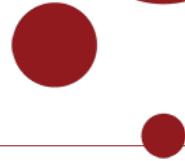


UNIVERSITY OF COPENHAGEN

DEPARTMENT OF ECONOMICS



## **Post-2021 Forecasting of Electricity Prices in Western Denmark: A Weather-Based Approach**

**Seminar: Asset Prices and Financial Markets**

**Alon Clausen and Johan Ølgaard**

**Advisor:** Jakob Lunbeck Serup

**Handed in:** December 20<sup>th</sup>, 2024

**Keystrokes** 45.000

**ECTS:** 15 points

[This page is intentionally left blank]

**Post-2021 Forecasting of Electricity Prices in Western Denmark:  
A Weather-Based Approach**

Alon Clausen<sup>†</sup> and Johan Ølgaard<sup>‡</sup>

Advisor: Jacob Lunbeck Serup

December 2024

University of Copenhagen  
Department of Economics

## Abstract

Electricity price forecasting has become increasingly difficult since 2021 due to heightened volatility driven by geopolitical instability and the rise of renewable energy. This study assesses the performance of SARIMA, SARIMAX, and LSTM models in forecasting day-ahead electricity prices in Western Denmark (DK1) using weather data and commodity prices. While all models outperform a naïve benchmark, SARIMAX achieves the lowest RMSE (313.63), and SARIMA has the lowest MAE (222.44). However, all models struggle to predict extreme price spikes above DKK 1000 per MWh, highlighting their limitations in volatile markets. We find more advanced forecasting techniques and the inclusion of richer datasets are necessary to encapsulate today's electricity price dynamics.

AI tools have been used for this paper solely for coding purposes.

### Responsibilities:

Jointly: 1, 5, and 6

Alon Clausen: 2.1.1, 3, 3.1, and 4.1

Johan Ølgaard: 2, 3.2 ,4, and 4.2

---

<sup>†</sup>Student ID: smr136

<sup>‡</sup>Student ID: jlh601

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Literature review . . . . .	2
<b>2</b>	<b>Data</b>	<b>4</b>
2.1	Electricity Prices . . . . .	4
2.1.1	Auto correlation in the price data . . . . .	5
2.2	Climate Data . . . . .	6
2.3	Commodity Prices . . . . .	8
2.4	Feature Engineering . . . . .	9
2.4.1	Principal Component Analysis . . . . .	9
2.4.2	Cyclical Encoding . . . . .	9
<b>3</b>	<b>Econometric Theory</b>	<b>10</b>
3.1	Classical Econometric Models . . . . .	10
3.1.1	SARIMA . . . . .	10
3.1.2	SARIMAX . . . . .	11
3.2	LSTM Neural Network . . . . .	11
3.2.1	Loss function regularisation . . . . .	13
3.2.2	Optimisation of LSTM NN using Adam . . . . .	14
<b>4</b>	<b>Empirical Analysis</b>	<b>15</b>
4.1	Classical Econometric Models . . . . .	17
4.2	LSTM Neural Network . . . . .	17
<b>5</b>	<b>Discussion</b>	<b>19</b>
<b>6</b>	<b>Conclusion</b>	<b>20</b>
	<b>Bibliography</b>	<b>22</b>
<b>A</b>	<b>Methods</b>	<b>26</b>
A.1	Principal Component Analysis . . . . .	26
A.2	Assessing predictive performance . . . . .	26
A.3	Diebold Mariano Test . . . . .	27
<b>B</b>	<b>Estimation</b>	<b>27</b>

# 1 Introduction

Since 2021, electricity price volatility has surged considerably, pushing prices to unprecedented levels. As demand for renewable energy rises, effective forecasting becomes essential for long-term investments in power grid infrastructure. The heightened geopolitical uncertainty following the onset of the Russo-Ukrainian war, coupled with the EU's efforts to reduce dependence on Russian fossil fuels, has increased prices and greater volatility in energy commodities.<sup>1</sup> With this in mind, it seems unlikely that electricity prices will return to previous levels or be less volatile in the near future. However, most recent academic papers within the field of electricity price forecasting still focus on predicting prices before 2021, hence in periods with much lower volatility, making their findings less applicable to the current situation.

This paper focuses on predicting day-ahead electricity prices in Western Denmark using weather data in combination with oil and gas prices using data up until November 29<sup>th</sup> 2024. Leveraging classical econometric models like the Seasonal AutoRegressive Integrated Moving Average with eXogenous variables (SARIMAX) as well as a state-of-the-art Long Short-Term Memory (LSTM) neural network, we examine if these models are capable of predicting the day-ahead electricity price.

We find that all the models outperform our naïve benchmark. Among these, the SARIMAX model performs best in terms of RMSE (313.63), while the SARIMA model achieves the lowest MAE (222.44). This suggests that the SARIMAX model is better at capturing sudden spikes in the data.

However, as eluded by the RMSE and MAE, all the models have difficulty predicting the day-ahead prices in this volatile environment, generally failing to predict prices much above DKK 1000 per MWh.

## 1.1 Literature review

The literature on electricity price forecasting has expanded significantly, covering both traditional econometric models and modern machine-learning approaches. It includes well-established methods such as various ARIMA<sup>2</sup> and GARCH<sup>3</sup> models, as well as more computationally demanding techniques like neural networks and hybrid frameworks.<sup>4</sup> To establish a foundation for our analysis, we review the studies relevant to our research objectives.

The ARIMA model is frequently used either as a baseline model or foundational component for more advanced models. Garcia et al. (2005) employ an ARIMA model to capture the trend and seasonal components of electricity prices and extend it with GARCH components to model the time-varying volatility they claim to be present in the data.

---

<sup>1</sup>Ferrari Minesso, Lappe, and Rößler 2024.

<sup>2</sup>Autoregressive Integrated Moving Average

<sup>3</sup>Generalized Autoregressive Conditional Heteroskedasticity

<sup>4</sup>Lago et al. 2020.

While ARIMA and GARCH models are foundational tools for forecasting, limitations have been identified when applying these models to energy markets. Escribano, Pena, and Villaplana (2002) highlights that electricity price spikes, which often act as outliers, can bias the estimation of GARCH processes, potentially leading to an integrated volatility process. Such a process suggests non-reverting volatility, which does not align with the observed mean-reverting behaviour that equilibrium electricity prices have demonstrated in the past.

To address these shortcomings, more recent research has shifted toward hybrid models and machine learning approaches. Karabiber and Xydis (2019) use an Artificial Neural Network (ANN) to forecast the DK1 electricity prices. These forecasts are compared against an ARIMA model with exogenous variables and a TBATS.<sup>5</sup> They show that the ARIMA model outperforms the other models when considering the mean of the daily mean absolute error (MAE).

Electricity prices have become significantly more volatile in recent years due to the instability in natural gas supplies following Russia's invasion of Ukraine.<sup>6</sup> Trebbien et al. (2023) introduce a Long Short-Term Memory (LSTM) model to forecast day-ahead electricity prices in the German-Luxembourg market in the years of 2019 to 2022. Their findings reveal a sharp decline in the model's predictive accuracy in 2022 compared to earlier years, with the MAE increasing almost tenfold. This decline highlights the growing volatility of the electricity market, underscoring the challenges forecasting models face in capturing such volatile dynamics.

As Europe transitions from reliance on Russian gas to a more diversified mix of energy sources, electricity price forecasting models must adapt to the increasing complexity of the energy market, especially as renewable sources are on the rise and production from these sources is highly variable and dependent on external conditions. One effective approach is to incorporate a wide array of variables, including power market indicators, commodity prices, and weather data. For example, Madadkhani and Ikonnikova (2024) use a machine learning setup with 71 distinct variables, showcasing the potential of advanced data-driven approaches to capture the intricacies of modern electricity markets. They find that the main driver in their model is the residual power load, energy demanded from non-renewable energy sources, which depends directly on the energy produced by weather-sensitive renewables.

We build upon the work of Trebbien et al. (2023), adopting their approach to electricity price forecasting, and take inspiration from Madadkhani and Ikonnikova (2024) by expanding the set of input variables. Specifically, we add to the existing literature by utilising more recent price data than any published paper and incorporating commodity prices

---

<sup>5</sup>Trigonometric Seasonal Box-Cox Transformation with ARMA residuals, Trend, and Seasonal Components

<sup>6</sup>Adolfsen et al. 2022.

and weather data. With this approach, we aim to address the growing complexity and increasing difficulty in forecasting electricity markets in the wake of shifting energy dynamics.

The remainder of this paper is organised as follows: In section 2, we present the data utilised in this study. In section 3, we outline the econometric theory that underpins our models. section 4 focuses on analysing the optimal parameters for our models and presents our findings. In section 5, we discuss the implications of our results, as well as the limitations of our model and the data available. Finally, in section 6, we conclude upon our findings.

## 2 Data

To conduct our analysis, we rely on data from three primary sources: the Danish Meteorological Institute (DMI), the Danish TSO Energinet and Yahoo Finance. The DMI and Energinet are both independent public enterprises owned by the Danish Ministry of Climate, Energy, and Utilities and provide access to their data through their APIs. Aggregated hourly DMI climate data is available from January 1<sup>st</sup>, 2011 and onwards,<sup>7</sup> while hourly electricity prices are available back to July 1<sup>st</sup>, 1999.<sup>8</sup> In addition, we pull daily commodity prices for Brent crude oil and TTF natural gas from Yahoo Finance. The TTF natural gas time series is only available from October 23<sup>rd</sup> 2017 and onwards.

Hence, we will use October 23<sup>rd</sup> 2017 as the starting point for our analysis - resulting in 62,303 hourly observations from this point until November 29<sup>th</sup> 2024. For evaluation purposes, we exclude the last 4 months,  $\sim 5\%$  of the total, from the training data; thus, all models are trained on data from October 23<sup>rd</sup> 2017 until July 31<sup>st</sup> 2024.

### 2.1 Electricity Prices

Energinet owns, operates, and develops the transmission systems for electricity and gas in Denmark. Together with other European TSOs, they own Nord Pool, where electricity trading in the day-ahead and intraday markets happen,<sup>9</sup> for both the DK1, Western Denmark, and DK2, Eastern Denmark,<sup>10</sup> areas based upon expected and actual production and demand. We will use the hourly day-ahead price in DK1 as the dependent variable for our analysis. Specifically, we use the dataset provided by Energi Data Service, Energinet (2024). In figure 2.1, we have plotted the full series.

---

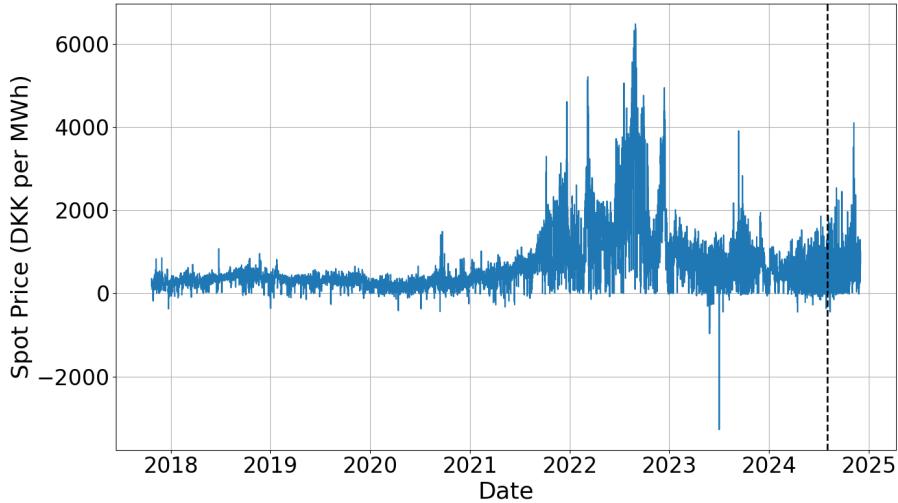
<sup>7</sup>Danish Meteorological Institute 2024.

<sup>8</sup>Energinet 2024.

<sup>9</sup>Nord Pool 2024.

<sup>10</sup>Split down the Great Belt strait, see figure 2.3 for exact split of municipalities.

Figure 2.1: Hourly electricity prices in DKK in DK1, October 23<sup>rd</sup> 2017 - November 29<sup>th</sup> 2024



NOTE: Hourly day-ahead electricity prices in DKK for DK1 from Energinet (2024). July 29<sup>th</sup> 2024 marked with the dotted line

From figure 2.1, it is apparent that the prices were relatively stable until mid-2021 when prices and volatility increased dramatically. This, in part, is explained by increases in both crude oil as well as natural gas<sup>11</sup> where especially natural gas prices have had a strong influence on electricity prices.<sup>12</sup>

In 2021, the EU began reducing gas imports from Russia amid rising global demand and declining European storage levels.<sup>13</sup> Further, the Russian invasion of Ukraine in the spring of 2022 strained commodity markets, keeping volatility high.<sup>14</sup>

### 2.1.1 Auto correlation in the price data

To further analyse the data and identify potential lags and seasonality, the autocorrelation function (ACF) is examined as seen in figure 2.2. The ACF reveals significant correlations at lags 24 and 48, indicating a strong daily seasonal pattern in the time series. The partial autocorrelation function (PACF) shows significant spikes in the first lag and smaller spikes at subsequent lags, suggesting the presence of short-term dependencies in the data. Furthermore, the lack of pronounced seasonal spikes in the PACF aligns with the seasonal pattern identified in the ACF, confirming the importance of including seasonal terms in the model.

---

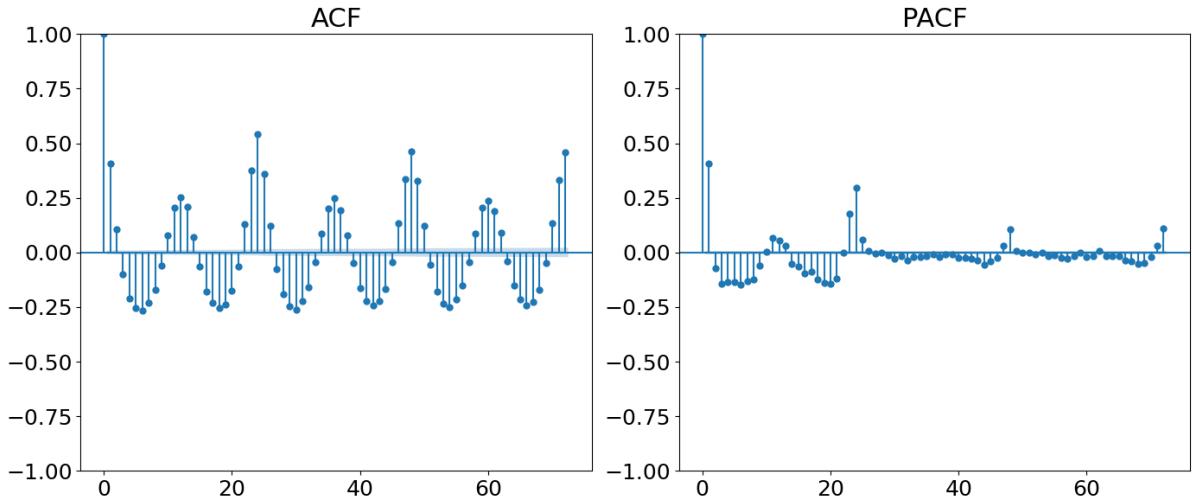
<sup>11</sup>See figures 2.6 and 2.7.

<sup>12</sup>Kuik et al. 2022.

<sup>13</sup>Fulwood 2022.

<sup>14</sup>Kuik et al. 2022.

Figure 2.2: ACF and PACF Plot



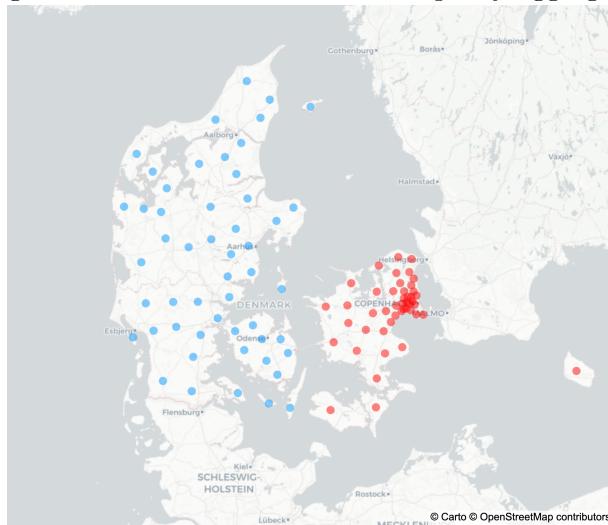
NOTE: Autocorrelation and Partial Autocorrelation for the electricity prices from Energinet (2024) shown in figure 2.1

Before introducing exogenous regressors such as temperature, wind speed, and wind direction, this analysis helps establish a baseline structure and guides the selection of the best-fitting models. These external variables are then incorporated to capture additional variations, further improving the predictive performance and interpretability of the model.

## 2.2 Climate Data

DMI collects data from over 200 measuring stations across Denmark and provides access to hourly aggregates through its Open Data program. We specifically use the Climate Data Danish Meteorological Institute (2024), which contains hour-old meteorological observations that have undergone quality control. The data we use is further aggregated on the municipality level, providing us with aggregated data points for every hour for every municipality in Denmark.

Figure 2.3: Centroid of each municipality aggregate

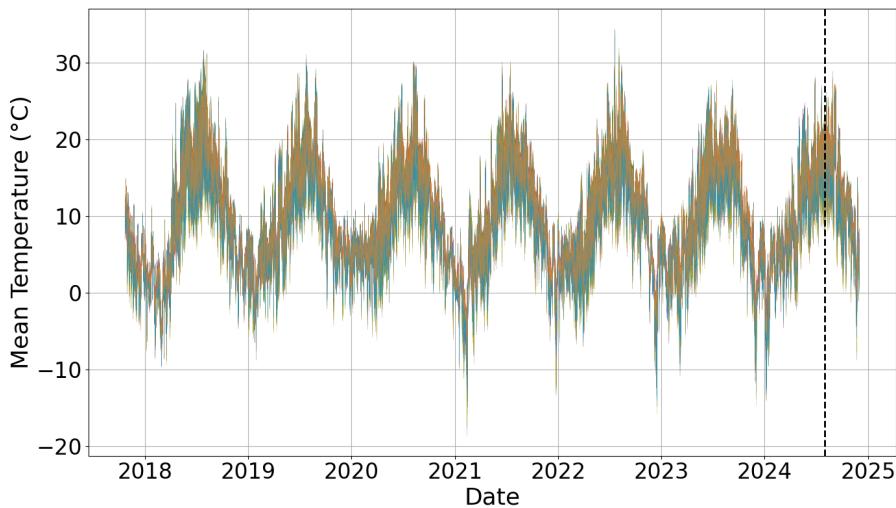


NOTE: Map is colour-coded with blue centroids being municipalities placed in the DK1 area and red municipalities in the DK2 area

As the electricity market in Denmark is split into DK1 and DK2, and our focus is the prices in DK1, we will solely consider the weather data in those municipalities.

DMI reports 19 different hourly measurements for each municipality; only ten have been consistently reported since 2017. In our analysis, we have chosen to focus on mean temperature, mean wind speed and mean wind direction, as we expect that wind speed and direction heavily influence the amount of renewable electricity produced by windmills, thus putting downward pressure on price while we expect the temperature to affect the demand-side, especially during colder and darker periods in winter where, demand for heating and lighting rises, leading to increased electricity consumption.

Figure 2.4: Hourly temperature, October 23<sup>rd</sup> 2017 - November 29<sup>th</sup> 2024

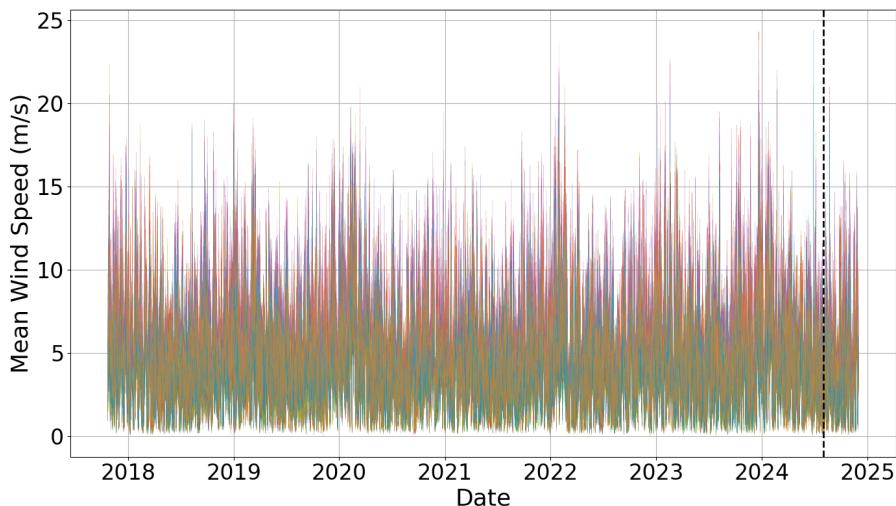


NOTE: Hourly temperature data for each municipality in DK1 from Danish Meteorological Institute (2024)

July 29<sup>th</sup> 2024 marked with the dotted line

As seen in figure 2.4 temperatures across Western Denmark are unsurprisingly highly correlated and seasonally dependent following a clear sinusoidal cycle.

Figure 2.5: Hourly wind speed, October 23<sup>rd</sup> 2017 - November 29<sup>th</sup> 2024



NOTE: Hourly wind speed data for each municipality in DK1 from Danish Meteorological Institute (2024)

July 29<sup>th</sup> 2024 marked with the dotted line

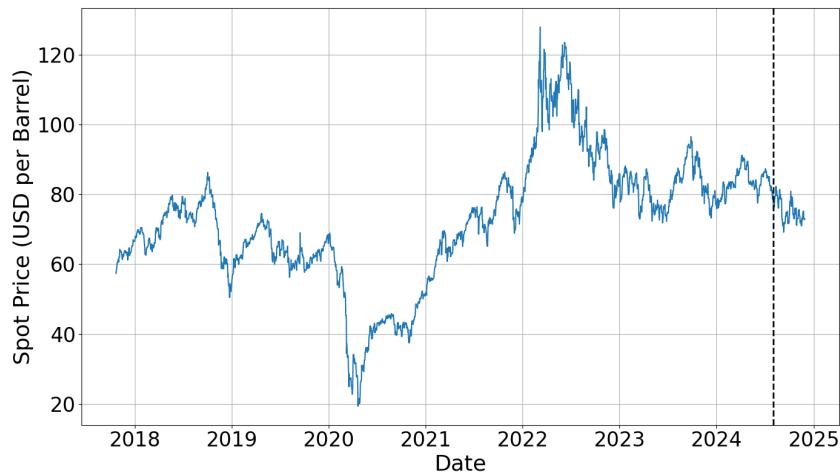
The windspeed data seen in figure 2.5 show that wind patterns are not as correlated nor seasonally dependent as the temperature illustrated in figure 2.4. However, there seems to be a slight trend with the windiest days tending to occur during the winter months, with occasional wind speeds reaching upwards of 20 meters per second.

## 2.3 Commodity Prices

We include prices of 2 additional commodities in our model: Brent crude oil and TTF natural gas.

Brent crude oil originates from the North Sea, making it a reliable benchmark for the European crude oil market. Furthermore, Brent benefits from high liquidity and extensive trading volumes in global markets, reinforcing its role as a key benchmark.<sup>15</sup> The time series is plotted in figure 2.6 below.

Figure 2.6: Daily Brent Crude Oil, October 23<sup>rd</sup> 2017 - November 29<sup>th</sup> 2024



NOTE: Daily Brent Crude Oil price in USD per Barrel from Yahoo Finance (2024a). July 29<sup>th</sup> 2024  
marked with the dotted line

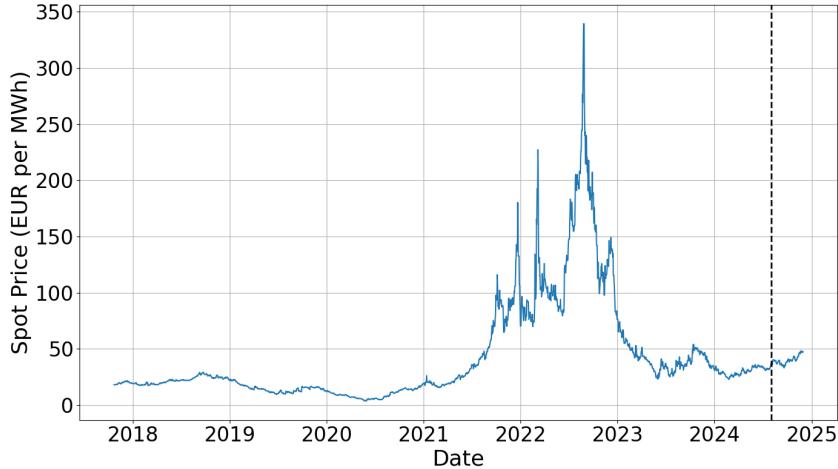
As seen in figure 2.6, the price of Brent Crude oil has seen pronounced swings from 2017 - 2024. Most notably, it dropped down, hitting prices around USD 20 per barrel in 2020 during COVID-19, its lowest level in over 20 years.<sup>16</sup> Just 2 years later, at the beginning of the most recent onset of the Russo-Ukrainian war, prices had increased sixfold to topping above USD 120 per barrel. Since 2023, prices have stabilised at around USD 80 per barrel.

The Title Transfer Facility (TTF) is the Dutch market for LNG managed by Gasunie Transport Services (2024) and Europe's largest market. TTF has a high liquidity and global supply of LNG, frequently serving as a benchmark for European LNG. The time series is plotted in figure 2.7 below.

<sup>15</sup>Trading Economics 2024.

<sup>16</sup>Gaffen 2022.

Figure 2.7: Daily TTF Natural Gas, October 23<sup>rd</sup> 2017 - November 29<sup>th</sup> 2024



NOTE: Daily TTF LNG price in EUR per MWh equivalents from Yahoo Finance (2024b). July 29<sup>th</sup> 2024 marked with the dotted line

As mentioned in section 2.1, gas storage levels and political uncertainty contributed to price increases throughout 2021. Following the Russian invasion of Ukraine in the spring of 2022, additional pressure was applied to natural gas prices, which again surged after the sabotage of the Nord Stream pipeline, causing TTF natural gas prices to reach record highs, particularly due to heightened supply risks.<sup>17</sup> However, over the past two years, prices have stabilised, albeit at a higher level seen before 2021.

## 2.4 Feature Engineering

### 2.4.1 Principal Component Analysis

As mentioned in section 2.2 especially, the temperature data seem highly correlated, while the wind data is more noisy. This can lead to multicollinearity, causing instability in coefficient estimates and increasing the risk of overfitting our models. Furthermore, as we have climate data for every municipality in the DK1 area, the data spans +200 variables, which computationally poses additional difficulties when running models locally.

To overcome this, we employ Principal Component Analysis (PCA) to reduce dimensionality, thus transforming the original set of correlated variables into a smaller set of uncorrelated principal components while keeping the majority of the variability in the data. We use the implementation by scikit-learn (2024) following Jolliffe (2002) as explained in section A.1 keeping at least 99% of the total variation in the temperature and in the wind data. Doing so enables us to reduce the number of explanatory variables to below 60.

### 2.4.2 Cyclical Encoding

To account for seasonality in electricity prices in all our models, we apply cyclical encoding to capture daily, weekly, and yearly patterns while preserving the circular nature of these variables. This approach has been shown to outperform traditional one-hot encoding in

---

<sup>17</sup>Goodell et al. 2023.

modelling seasonal features.<sup>18</sup> The cyclical nature is represented using the following sine and cosine transformations for each observation:

$$\begin{aligned} \text{hour\_sin} &= \sin\left(2\pi \cdot \frac{\text{time of day}}{24}\right), & \text{hour\_cos} &= \cos\left(2\pi \cdot \frac{\text{time of day}}{24}\right) \\ \text{week\_sin} &= \sin\left(2\pi \cdot \frac{\text{day of week}}{7}\right), & \text{week\_cos} &= \cos\left(2\pi \cdot \frac{\text{day of week}}{7}\right) \\ \text{month\_sin} &= \sin\left(2\pi \cdot \frac{\text{month}}{12}\right), & \text{month\_cos} &= \cos\left(2\pi \cdot \frac{\text{month}}{12}\right) \end{aligned} \quad (2.1)$$

We encode the time of day, day of the week, and month, as human behaviour heavily influences electricity consumption and prices. Demand is expected to be highest during peak load hours on workdays, and due to limited cooling usage during the summer, demand is expected to be higher during the winter.

In addition to the cyclical encoding of time variables, we also encode the mean wind direction cyclically in a similar fashion.

$$\text{wind\_dir\_sin} = \sin\left(2\pi \cdot \frac{\text{wind\_dir} \cdot \pi}{180}\right) \quad \text{wind\_dir\_cos} = \cos\left(2\pi \cdot \frac{\text{wind\_dir} \cdot \pi}{180}\right) \quad (2.2)$$

This allows us to accurately model the relative proximity of different wind directions, ensuring smooth transitions between angles.

### 3 Econometric Theory

We employ the classical econometric SARIMA and SARIMAX models and a state-of-the-art LSTM neural network to forecast day-ahead electricity spot prices.

#### 3.1 Classical Econometric Models

The SARIMA and SARIMAX models build upon the classical ARIMA framework by incorporating seasonality. Additionally, the SARIMAX model includes the option for exogenous variables. This section presents both extensions.

##### 3.1.1 SARIMA

To model time series data exhibiting seasonality, a standard ARIMA model is extended to include seasonal components, resulting in the Seasonal ARIMA (SARIMA) model. The SARIMA model incorporates both non-seasonal and seasonal factors in the autoregressive and moving average parts, as well as seasonal differencing to achieve stationarity. The SARIMA model is denoted as  $\text{SARIMA}(p, d, q)(P, D, Q)_s$ , where  $p$  is the order of the autoregressive part,  $d$  is the number of differences required for stationarity,  $q$  is the order of the moving average part.  $P, D, Q$  are equivalent but just for the seasonal part.  $s$  is the length of the seasonal period. We use the backshift notation presented by Hyndman and

---

<sup>18</sup>Mahajan, Singh, and Bruns 2021.

Athanassopoulos (2018). The general form of the SARIMA model is:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^Dy_t = c + \Theta(B^s)\theta(B)\varepsilon_t \quad (3.1)$$

Where  $\phi(B)$  and  $\theta(B)$  are the non-seasonal AR and MA polynomials,  $\Phi(B^s)$  and  $\Theta(B^s)$  are the seasonal AR and MA polynomials,  $(1 - B^s)^D$  represents the seasonal differencing operator applied  $D$  times, and  $\varepsilon_t$  is the error term at time  $t$ . The seasonal polynomials are defined as:

$$\begin{aligned}\phi(B) &= 1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p \\ \theta(B) &= 1 + \theta_1B + \theta_2B^2 + \cdots + \theta_qB^q \\ \Phi(B^s) &= 1 - \Phi_1B^s - \Phi_2B^{2s} - \cdots - \Phi_PB^{Ps} \\ \Theta(B^s) &= 1 + \Theta_1B^s + \Theta_2B^{2s} + \cdots + \Theta_QB^{Qs}\end{aligned}$$

This formulation allows the model to capture both non-seasonal and seasonal behaviours in the time series data.<sup>19</sup>

### 3.1.2 SARIMAX

To incorporate external influences into the model, such as weather variables and commodity prices, we extend the SARIMA model to include exogenous variables, resulting in the Seasonal ARIMA with exogenous variables (SARIMAX) model. The SARIMAX model can be expressed as:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^Dy_t = c + \beta X_t + \Theta(B^s)\theta(B)\varepsilon_t \quad (3.2)$$

Where  $X_t$  is a vector of exogenous variables at time  $t$  and  $\beta$  is a vector of coefficients corresponding to the exogenous variables. The other terms are as previously defined. Including exogenous variables allows the model to account for external factors that can influence the dependent variable  $y_t$ , improving the accuracy of forecasts.<sup>20</sup>

## 3.2 LSTM Neural Network

As our model predicts prices into the future using the weather of the past, the model has to know which way the weather is trending to best predict where prices are trending. Hence, we utilise a Long Short-Term Memory (LSTM) Neural Network, which falls under the classification of recurrent neural network (RNN). The LSTM model was first developed by Hochreiter and Schmidhuber (1997) and is designed to model sequential data by maintaining a memory of previous inputs.

For the *forward pass*, we rely on the implementation of the model presented in Sak,

---

<sup>19</sup>Alharbi and Csala 2022.

<sup>20</sup>Ibid.

Senior, and Beaufays (2014) where the network computes a mapping from an input sequence  $x = (x_1, \dots, x_T)$ , of weather data and previous prices to an output sequence  $\hat{y} = (\hat{y}_1, \dots, \hat{y}_T)$  of future prices. The model predicts the output with the current model weights for each input sequence by passing the data through a set of gates and activations. These predicted values are then used to update the weights through a backward propagation loop.

Each hidden dimension in each hidden layer starts by computing the *input gate*,  $i_t$ , determining how much information should continue into its "memory."

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \quad (3.3)$$

Where  $\sigma(\cdot)$  is the sigmoid function of form  $\sigma(x) = \frac{1}{1+e^{-x}}$ ,  $W_{ix}$ ,  $W_{im}$ , and  $W_{ic}$  are weight matrices of the input at time  $t$ .  $m_{t-1}$  is the cell output activation vector from the previous time step, and  $c_{t-1}$  is the cell activation vector from the previous time step, while  $b_i$  is the input gate bias vector.

Next, the *forget gate* is calculated in a similar fashion and used to control what part from  $c_{t-1}$  should be "forgotten."

$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f) \quad (3.4)$$

Where  $W_{fx}$ ,  $W_{fm}$ , and  $W_{fc}$  are the weight matrices and  $b_f$  is the bias.

Next, the *cell activation vector* is calculated using (3.3) and (3.4).

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \quad (3.5)$$

Where  $\odot$  is the Hardmdard product,  $\tanh(\cdot)$  is the hyperbolic tangent function of form  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  and  $W_{cx}$  and  $W_{cm}$  are the weight matrices and  $b_c$  again is the bias. This cell state is the core part of the LSTM model, which carries information across states. It allows the model to selectively modulate information at the feature level by adjusting weights. This enables the model to adapt to new inputs, retain patterns, and suppress noise, thus retaining its memory across periods.

This is used to calculate the *output gate* determining what part of the cell's memory influences output.

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o) \quad (3.6)$$

Where  $W_{ox}$ ,  $W_{om}$ , and  $W_{oc}$  are the weight matrices and  $b_o$  is the bias. Notably, we here use the same-period cell activation vector from (3.5),  $c_t$  and not the cell activation vector from last period,  $c_{t-1}$  as used in (3.3) and (3.4).

Next, the *cell output activation vector* is calculated using (3.5) and (3.6).

$$m_t = o_t \odot \tanh(c_t) \quad (3.7)$$

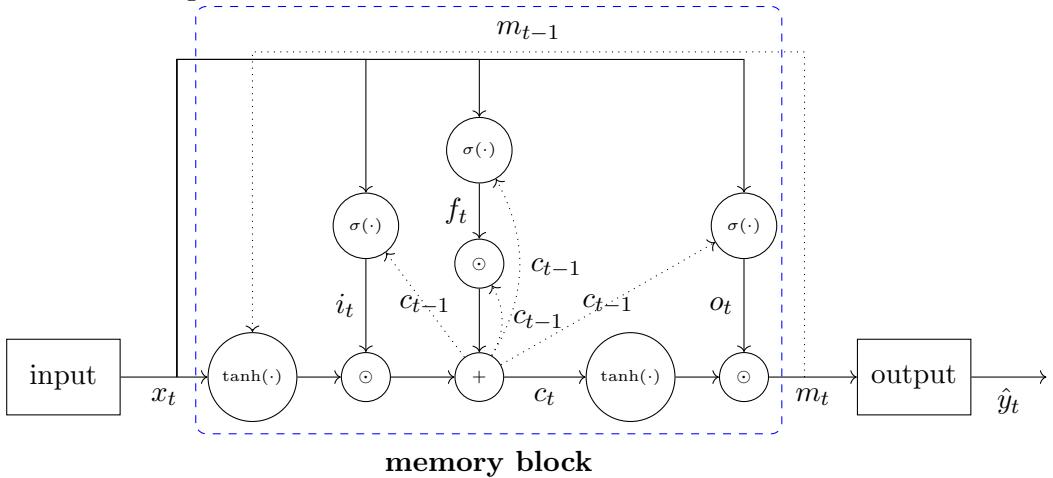
Lastly, we can calculate the *predicted* output sequence using (3.7).

$$\hat{y}_t = W_{ym}m_t + b_y \quad (3.8)$$

Where  $W_{ym}$  is the weight matrix for the output activation vector, and  $b_y$  is the bias for the output.

For each hidden layer in each hidden model dimension, this dynamic is computed in the *memory block* for every sequence. The flow of information in one hidden layer is illustrated in figure 3.1 below. In practice, the output function handles input from multiple memory blocks in every hidden layer.

Figure 3.1: Architecture in a node in an LSTM-based RNN



NOTE: Architecture of a single node LSTM-based RNN with a recurrent layer. Dotted line indicates value from previous period

From figure 3.1, we see how the cell output activation vector from time  $t - 1$  flows back into the memory block, preserving knowledge of previous data. This memory block adds an autoregressive-like memory component to the neural network.

### 3.2.1 Loss function regularisation

Though we have used PCA to reduce the number of variables in our model, it still seems reasonable to believe that not all of the remaining  $\sim 40$  wind directions variables are important parameters for the model. We, therefore, want to encourage sparsity in our LSTM RNN to reduce the likelihood of the model overfitting to the training data and staying more generalised. To do so, we utilise L1, or Lasso, regularisation as introduced by Tibshirani (1996). The Lasso adds the sum of the absolute value of the coefficients to the loss function, thereby encouraging sparsity. To combat overfitting further, we also implement L2 regularisation, or Ridge Regression, following Google (2024). Here, we add the sum of the squared weights to the loss function to penalise large weights in the network. This encourages the model to maintain smaller, more generalised weights.

Hence, the loss function when using Lasso and ridge regularisation and mean squared

error is:

$$\ell_t = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 + \lambda_{L1} \sum_{k \in K} |w_k| + \lambda_{L2} \sum_{k \in K} w_k^2 \quad (3.9)$$

Where  $w_k$  is the weights of the  $k^{\text{th}}$  explanatory variable, while  $\lambda_{L1}$  and  $\lambda_{L2}$  are the penalty parameters. As seen from (3.9), this loss function penalises the model for every coefficient different from 0, thereby minimising the number of explanatory variables in the model to the ones best at minimising the root mean square error term.

### 3.2.2 Optimisation of LSTM NN using Adam

We optimise using backpropagation through time (BPTT) with Adaptive Moment Estimation (Adam) as proposed by Kingma and Ba (2017), updating parameters for each batch to ultimately solve:

$$\theta^* = \arg \min_{\theta} \ell_t(\theta) \quad (3.10)$$

Where  $\theta$  are the parameters used for the last forward pass, i.e. all weight matrices and biases,  $W$  and  $b$  and  $\ell_t$  is the loss function specified in (3.9).

For the optimisation we initialise the weight matrices,  $W$ , using Xavier initialisation<sup>21</sup>, drawing all the weights from:

$$W \sim U \left( -\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}} \right) \quad (3.11)$$

Where  $U$  is the uniform distribution, and  $n$  is the size of the previous layer, all the biases,  $b$ , are set to 0. We then use Adam to update the weights by first computing the gradients of the loss functions around the previous parameters estimates,  $\theta_{t-1}$ .

$$g_t = \nabla_{\theta} \ell_t(\theta_{t-1}) \quad (3.12)$$

This is used to update the biased first- and second-moment estimates.

$$s_t = \beta_1 \cdot s_{t-1} + (1 - \beta_1) \cdot g_t \quad (3.13)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (3.14)$$

Where  $s_t$  is the first-moment estimate,  $v_t$  is the second-moment estimate, and  $\beta_1$  and  $\beta_2$  are the exponential decay rates defining how much should be kept from the previous state and how much emphasis should be put on the new gradients. Next, we bias correct the

---

<sup>21</sup>Glorot and Bengio 2010.

moments with:

$$\hat{s}_t = \frac{s_t}{1 - \beta_1^t} \quad (3.15)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3.16)$$

The bias correction ensures that the initial moments,  $m_0 = 0$  and  $v_0 = 0$ , do not influence the final parameters too much. Lastly, this is used to update the parameters.

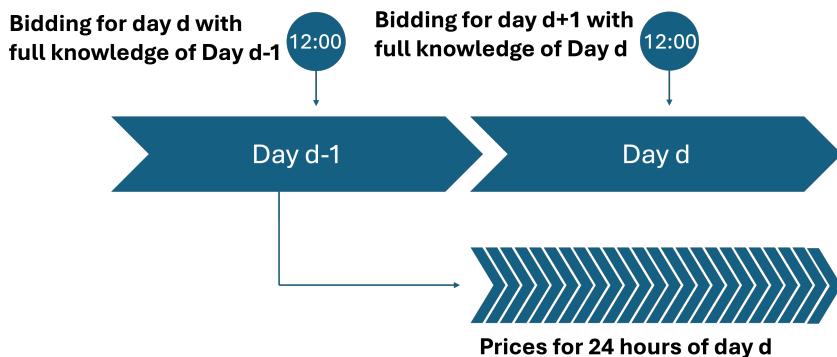
$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{\hat{s}_t}{\sqrt{\hat{v}_t + \varepsilon}} \quad (3.17)$$

Where  $\alpha$  is the learning rate and  $\varepsilon$  is a small constant to ensure numerical stability. The updated parameters are then used in the next forward pass as presented in (3.3)-(3.6), which is then used to update the parameters again.

## 4 Empirical Analysis

The day-ahead market closes at 12:00 the day prior. Hence, to predict the price for every hour in the day-ahead market in day  $d$ , we use the data available up until 12:00 on  $d - 1$ , including the full day-ahead spot price from  $d - 1$  as illustrated in figure 4.1.

Figure 4.1: Illustration of the day-ahead auction market



NOTE: To predict the day-ahead market, we use the information available up until market closing on Nordpool at 12:00 the prior day

As mentioned in section 2, we use data up until July 31<sup>st</sup> 2024 as training, while performance is evaluated based on the predicting power from August 1<sup>st</sup> 2024 for all models. Our results are presented in table 4.1, where we report a naïve benchmark computed as simply assuming prices will be equal to the prices the day prior.

Table 4.1: Performance of predictions for results for the

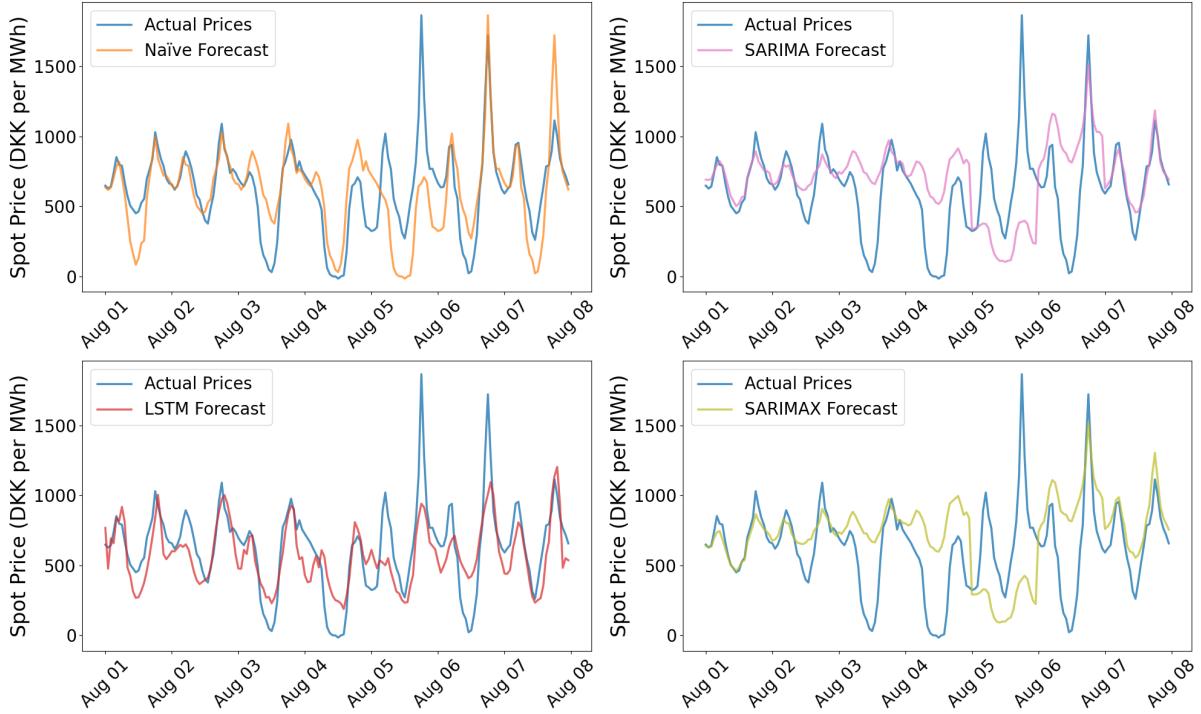
	Naïve	SARIMA	SARIMAX	LSTM
RMSE	354.00	318.47 (0.037)	313.63 (0.011)	324.04 (0.133)
MAE	253.95	222.44 (0.013)	224.69 (0.013)	249.27 (0.717)

NOTE: Performance calculated as the root mean square error and the mean absolute error as specified in section A.2. Diebold-Mariano test  $p$ -valuem as specified in section A.3 with the naïve as benchmark, in parenthesis

From table 4.1, we see all our models outperform the naïve benchmark; however, neither truly manages to capture the market dynamics. When performing the Diebold-Mariano test, we can only reject the null against the naïve prediction with the SARIMA and SARIMAX models.

The LSTM is performing the worst of our models, as it struggles to capture the sudden volatility experienced in the data even if the daily spike persists for a few days. The SARIMA and SARIMAX do manage to capture part of this when prices are high for more than a day; however, especially during these extreme spikes, the naïve model performs best, as seen in figure B.3. In periods with low volatility, the LSTM model outperforms the others, as seen in figure 4.2.

Figure 4.2: Forecasts for August 1<sup>st</sup> 2024 until August 8<sup>th</sup> 2024



NOTE: Forecast for a period with low volatility. The LSTM model outperforms the other models

## 4.1 Classical Econometric Models

The SARIMA and SARIMAX presented in section 3.1 are fitted using the `auto.arima` function from the `pmdarima` library<sup>22</sup>, an implementation based on the algorithm proposed by Hyndman and Athanasopoulos (2018). This approach selects parameters that minimise the Akaike Information Criterion (AIC). The models are fitted on the dataset from October 23<sup>rd</sup> 2017 until July 31<sup>st</sup>, 2024. The models generate daily forecasts of the day-ahead spot price by using all available historical data up to the forecasted day. This is done to replicate that when predicting prices for day  $d$ , all the prices from day  $d - 1$  are available 4.1. To address overfitting, we apply Lasso regularisation to both models. This corresponds to using the loss function in equation (3.9) with the  $\lambda_2$ -term set to zero, leaving only the  $L_1$ -penalty. We apply the Lasso using 5-fold cross-validation to the cyclically encoded date components in the SARIMA model and all exogenous variables in the SARIMAX model.

**SARIMA** By examining the ACF and PACF plots in figure 2.2, we identify daily seasonality, which serves as the basis for selecting the best fitting SARIMA model where the daily seasonality results in  $s=24$ . This SARIMA model only includes daily seasonality, so we use the cyclically encoded date components. We find that the best-fitted model is given by

$$\Phi(B^{24})\phi(B)y_t = c + \Theta(B^{24})\epsilon_t \quad (4.1)$$

Where

$$\begin{aligned} \Phi(B^{24}) &= 1 - \Phi_1 B^{24} - \Phi_2 B^{48} \\ \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 \\ \Theta(B^{24}) &= 1 \end{aligned} \quad (4.2)$$

**SARIMAX** The weather data and commodity prices are included as exogenous variables. By including these variables, we find that the best-fitted model has the same seasonal and non-seasonal AR and MA terms as the SARIMA but now includes a vector for the exogenous variables. Therefore, the best-fitted SARIMAX model is given by:

$$\Phi(B^{24})\phi(B)y_t = c + \beta X_t + \Theta(B^{24})\epsilon_t, \quad (4.3)$$

Where the polynomials are identical to those in the SARIMA specification (4.2), and  $X_t$  is the vector of all exogenous variables.

## 4.2 LSTM Neural Network

For training the LSTM model described in section 3.2, we split the training data into two sets: training and validation in addition to the test dataset already described in section 2. The model trains on the data from October 23<sup>rd</sup> 2017 until July 31<sup>st</sup> 2023 and validates

---

<sup>22</sup>Taylor G. Smith 2021.

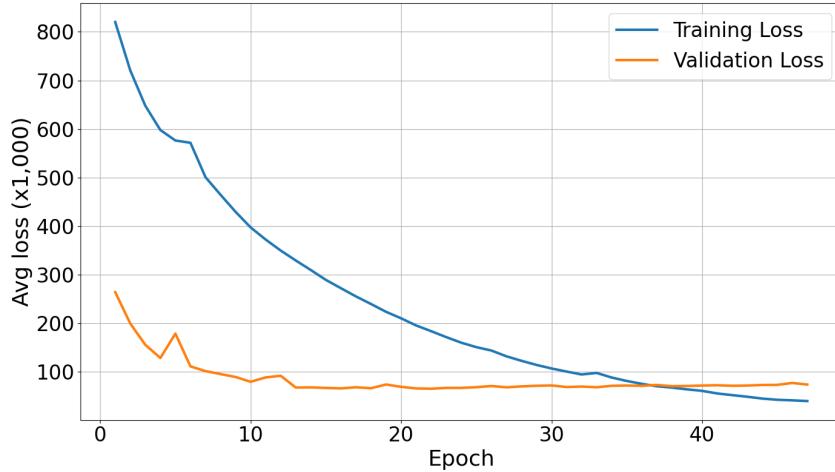
on data from August 1st 2023 until July 31st 2024 to ensure it is not overfitting. The model performance is evaluated using (3.9) for every sequence and passed backwards using BPTT as described in section 3.2.2.

To determine optimal hyperparameter values, we have conducted a grid search across the parameter values specified in table B.1. We have found the optimal hyperparameter specification to be

$$\begin{aligned}
 \lambda_1^* &= 10^{-1} \\
 \lambda_2^* &= 10^{-4} \\
 \text{hidden dimensions}^* &= 512 \\
 \text{hidden layers}^* &= 2 \\
 \text{batch size}^* &= 256 \\
 \text{learning rate}^* &= 10^{-3}
 \end{aligned} \tag{4.4}$$

Using the parameter specification shown in (4.4), we run the LSTM neural network, seeing great improvement in the validation loss for the first 13 periods where after improvement over each epoch, declines as seen from figure 4.3. The best-performing model on the out-of-sample validation data is found after 22 epochs.

Figure 4.3: Model loss in each epoch



NOTE: Training loss is as defined in (3.9) while the validation loss is based solely on mean square error, i.e.

$$(3.9) \text{ without regularisation terms}$$

As seen in table 4.1, our LSTM model outperforms the naïve model. However, it struggles with capturing the spikes in electricity prices as seen in figure B.3 and hence is outperformed by the classical econometric SARIMA and SARIMAX models. However, it conversely outperforms the other models during less volatile times, as exemplified in figure 4.2.

## 5 Discussion

### Forecasting Challenges in Volatile Energy Markets

Our empirical findings show the significant challenges of forecasting electricity prices in today's volatile market environment. While all our models outperform the naïve benchmark, their predictive accuracy is notably worse than the results reported in earlier studies. For example, Karabiber and Xydis (2019) achieved an MAE of 33.24 when forecasting the DK1 market for 212 days in 2017 with similar approaches as ours.

Similarly, other pre-2021 studies reported low error metrics.<sup>23</sup> In contrast, Trebbien et al. (2023) demonstrated a sharp decline in accuracy between 2019 and 2022, with their model's MAE increasing from 27.82 in 2019 to 222.65 in 2022<sup>24</sup> – a value roughly comparable to ours, despite being evaluated on different datasets.

### Factors Driving Market Volatility

This decline in performance highlights the impact of increased price volatility in recent years. Further, geopolitical instability, particularly the Russo-Ukrainian war, has disrupted global energy markets, leading to supply chain challenges and increased volatility. Additionally, the growing share of renewable energy sources introduces variability due to their intermittent nature,<sup>25</sup> which has led to an increased focus on the residual load.<sup>26</sup>

### Data and Modeling Constraints

The complexity and interdependence of variables influencing electricity prices present challenges for forecasting models. Incorporating additional data could potentially enhance accuracy. This could include relevant trading data from neighbouring markets, e.g., DK2, DE/LU, SE4, and NO2 - where all but SE4 are directly connected to DK1, intraday price data capturing day-ahead inaccuracies, and futures prices providing valuable forward-looking insights. Similarly, replacing historical weather data with real-time forecasts would better account for the variability in renewable energy generation, which depends heavily on meteorological factors, as our model is limited by solely having historical weather.

Both the SARIMA and SARIMAX models exhibit large error terms, reflecting their limitations in capturing the volatility and complexity of the current electricity market. While the inclusion of a GARCH term could potentially improve their performance by accounting for volatility clustering, the irregular and non-time-dependent nature of price spikes in the day-ahead market complicates its application. Moreover, SARIMAX's fixed coefficients for exogenous variables restrict its ability to adapt dynamically to changing market relationships, further reducing its effectiveness in volatile conditions.

In contrast, the LSTM model offers greater flexibility due to its ability to dynam-

---

<sup>23</sup>Lago et al. 2020; Garcia et al. 2005.

<sup>24</sup>The MAE values have been converted from EUR to DKK

<sup>25</sup>Eurostat 2024.

<sup>26</sup>Madadkhani and Ikonnikova 2024.

ically adjust to patterns through its input and output gates. However, its forecasting performance is highly dependent on the quality and relevance of its training data. The LSTM model requires sufficient data from market conditions that mirror the current highly volatile environment to capture the dynamics of extreme price fluctuations. Stressing the importance of richer datasets, such as those mentioned earlier, to improve the performance of both classical and advanced models.

### Future of electricity price forecasting

Our study suggests that models previously considered strong forecasters are now falling short in accuracy. This decline is most likely due to changes in electricity supply dynamics in Northern Europe, where prices now depend on a much broader array of variables. To address this increased complexity and volatility, an option could be the use of temporal fusion transformer neural networks.<sup>27</sup>

TFTs are designed to handle high-dimensional time series data, making them well-suited for forecasting environments with many interdependent variables, such as the electricity market. By leveraging more advanced attention mechanisms to focus on relevant features dynamically, the model can better adapt to both short-term fluctuations and long-term trends. Additionally, they are easier to interpret than other neural networks, allowing market participants to understand which factors drive the predictions.

Future research could explore how this method performs compared to traditional approaches like SARIMA or other neural networks like the LSTM in highly volatile markets. Furthermore, incorporating additional data sources into these advanced models could provide a more comprehensive understanding of electricity price dynamics.

## 6 Conclusion

In this paper, we examined the predictive performance of SARIMA, SARIMAX, and LSTM models in forecasting day-ahead electricity prices in Western Denmark’s volatile post-2021 market environment (DK1). Using weather data and commodity prices, we investigate the capacity of these models to capture these prices.

We find that all the models outperform the naïve benchmark, with the SARIMAX model achieving the lowest RMSE (313.63) and the SARIMA model yielding the lowest MAE (222.44). This suggests that the SARIMAX model is better at capturing sudden price spikes, while SARIMA provides slightly more accurate predictions on average; however, their scores are very similar.

As reflected in their error metrics, all our models struggle to predict extreme day-ahead prices, particularly when prices exceed DKK 1000 per MWh, highlighting the increasing difficulty of forecasting in today’s volatile energy markets. This suggests that models previously used in electricity price forecasting no longer hold the same predictive accuracy,

---

<sup>27</sup>Lim et al. 2020.

emphasising the need for advanced techniques and richer data to encapsulate market dynamics fully.

## Bibliography

- Adolfsen, Jakob Feveile et al. (June 2022). “The impact of the war in Ukraine on euro area energy markets”. en. In: URL: [https://www.ecb.europa.eu/press/economic-bulletin/focus/2022/html/ecb.ebbox202204\\_01~68ef3c3dc6.en.html](https://www.ecb.europa.eu/press/economic-bulletin/focus/2022/html/ecb.ebbox202204_01~68ef3c3dc6.en.html) (visited on 12/16/2024).
- Alharbi, Fahad Radhi and Denes Csala (Oct. 2022). “A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach”. en. In: *Inventions* 7.4, p. 94. ISSN: 2411-5134. DOI: 10.3390/inventions7040094. URL: <https://www.mdpi.com/2411-5134/7/4/94> (visited on 11/20/2024).
- Danish Meteorological Institute (2024). *DMI Open Data, Climate Data*. en. URL: [https://opendatadocs.dmi.govcloud.dk/en/Data/Climate\\_Data](https://opendatadocs.dmi.govcloud.dk/en/Data/Climate_Data) (visited on 11/10/2024).
- Diebold, Francis X. and Roberto S. Mariano (July 1995). “Comparing Predictive Accuracy”. In: *Journal of Business & Economic Statistics* 13.3. Publisher: [American Statistical Association, Taylor & Francis, Ltd.], pp. 134–144. ISSN: 0735-0015. URL: <https://www.jstor.org/stable/1392155> (visited on 12/18/2024).
- Energinet (Dec. 2024). *Elspot Price*. en. URL: <https://www.energidataservice.dk/tso-electricity/Elspotprices> (visited on 12/02/2024).
- Escribano, Álvaro, J. Ignacio Pena, and Pablo Villaplana (2002). “Modeling Electricity Prices: International Evidence”. en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.299360. URL: <http://www.ssrn.com/abstract=299360> (visited on 12/18/2024).
- Eurostat (2024). *Renewable energy statistics*. en. URL: [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Renewable\\_energy\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Renewable_energy_statistics) (visited on 12/17/2024).
- Ferrari Minesso, Massimo, Marie-Sophie Lappe, and Denise Rößler (Jan. 2024). “Geopolitical risk and oil prices”. en. In: *ECB Economic Bulletin* 2023.8. ISSN: xxxx-xxxx. URL: [https://www.ecb.europa.eu/press/economic-bulletin/focus/2024/html/ecb.ebbox202308\\_02~ed883ebf56.en.html](https://www.ecb.europa.eu/press/economic-bulletin/focus/2024/html/ecb.ebbox202308_02~ed883ebf56.en.html) (visited on 12/16/2024).
- Fulwood, Mike (Jan. 2022). *Surging 2021 European Gas Prices – Why and How?* en. Tech. rep. The Oxford Institute for Energy Studies, p. 8. URL: <https://www.oxfordenergy.org/wpcms/wp-content/uploads/2022/01/Surging-2021-European-Gas-Prices-%E2%80%93-Why-and-How.pdf> (visited on 12/01/2024).
- Gaffen, David (Feb. 2022). “Oil’s journey from worthless in the pandemic to \$100 a barrel”. en. In: *Reuters*. URL: <https://www.reuters.com/business/energy/oils-journey-worthless-pandemic-100-barrel-2022-02-24/> (visited on 12/14/2024).
- Garcia, R.C. et al. (May 2005). “A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices”. en. In: *IEEE Transactions on Power Systems* 20.2, pp. 867–874.

- ISSN: 0885-8950. DOI: 10.1109/TPWRS.2005.846044. URL: <http://ieeexplore.ieee.org/document/1425583/> (visited on 12/03/2024).
- Gasunie Transport Services (2024). *TTF*. en. URL: <https://www.gasunietransportservices.nl/en/about-gts/gas-transport/ttf> (visited on 12/12/2024).
- Glorot, Xavier and Yoshua Bengio (Mar. 2010). “Understanding the difficulty of training deep feedforward neural networks”. en. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. ISSN: 1938-7228. JMLR Workshop and Conference Proceedings, pp. 249–256. URL: <https://proceedings.mlr.press/v9/glorot10a.html> (visited on 12/01/2024).
- Goodell, John W. et al. (Sept. 2023). “Global energy supply risk: Evidence from the reactions of European natural gas futures to Nord Stream announcements”. en. In: *Energy Economics* 125, p. 106838. ISSN: 01409883. DOI: 10.1016/j.eneco.2023.106838. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0140988323003365> (visited on 12/12/2024).
- Google (2024). *Overfitting: L2 regularization — Machine Learning*. en. URL: <https://developers.google.com/machine-learning/crash-course/overfitting/regularization> (visited on 12/02/2024).
- Hochreiter, Sepp and Jürgen Schmidhuber (Nov. 1997). “Long Short-Term Memory”. In: *Neural Computation* 9.8, pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: <https://doi.org/10.1162/neco.1997.9.8.1735> (visited on 11/24/2024).
- Hoerl, Arthur E. and Robert W. Kennard (1970). “Ridge Regression: Biased Estimation for Nonorthogonal Problems”. In: *Technometrics* 12.1. Publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], pp. 55–67. ISSN: 0040-1706. DOI: 10.2307/1267351. URL: <https://www.jstor.org/stable/1267351> (visited on 12/02/2024).
- Hyndman, Rob J and George Athanasopoulos (May 2018). *Forecasting: Principles and Practice*. en. 2nd Edition. Monash University, Australia.
- Jolliffe, Ian (2002). *Principal Component Analysis*. en. Springer Series in Statistics. New York: Springer-Verlag. ISBN: 978-0-387-95442-4. DOI: 10.1007/b98835. URL: <http://link.springer.com/10.1007/b98835> (visited on 12/14/2024).
- Karabiber, Orhan Altuğ and George Xydis (Mar. 2019). “Electricity Price Forecasting in the Danish Day-Ahead Market Using the TBATS, ANN and ARIMA Methods”. en. In: *Energies* 12.5, p. 928. ISSN: 1996-1073. DOI: 10.3390/en12050928. URL: <https://www.mdpi.com/1996-1073/12/5/928> (visited on 10/21/2024).
- Kingma, Diederik P. and Jimmy Ba (Jan. 2017). *Adam: A Method for Stochastic Optimization*. arXiv:1412.6980. DOI: 10.48550/arXiv.1412.6980. URL: <http://arxiv.org/abs/1412.6980> (visited on 11/29/2024).
- Kuik, Friderike et al. (June 2022). “Energy price developments in and out of the COVID-19 pandemic – from commodity prices to consumer prices”. en. In: *ECB Economic Bulletin*

- 4/2022. URL: [https://www.ecb.europa.eu/press/economic-bulletin/articles/2022/html/ecb.ebart202204\\_01~7b32d31b29.en.html](https://www.ecb.europa.eu/press/economic-bulletin/articles/2022/html/ecb.ebart202204_01~7b32d31b29.en.html) (visited on 12/12/2024).
- Lago, Jesus et al. (Dec. 2020). *Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark*. arXiv:2008.08004. DOI: 10.48550/arXiv.2008.08004. URL: <http://arxiv.org/abs/2008.08004> (visited on 12/03/2024).
- Lim, Bryan et al. (Sept. 2020). *Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting*. arXiv:1912.09363 [stat]. DOI: 10.48550/arXiv.1912.09363. URL: <http://arxiv.org/abs/1912.09363> (visited on 12/15/2024).
- Madadkhani, Shiva and Svetlana Ikonnikova (Jan. 2024). “Toward high-resolution projection of electricity prices: A machine learning approach to quantifying the effects of high fuel and CO<sub>2</sub> prices”. en. In: *Energy Economics* 129, p. 107241. ISSN: 01409883. DOI: 10.1016/j.eneco.2023.107241. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0140988323007399> (visited on 12/16/2024).
- Mahajan, Tanisha, Gaurav Singh, and Glenn Bruns (Mar. 2021). “An Experimental Assessment of Treatments for Cyclical Data”. en. In: *Computer Science Conference for CSU Undergraduates*.
- Nord Pool (2024). *Nord Pool*. URL: <https://www.nordpoolgroup.com/en/> (visited on 11/29/2024).
- Ølgaard, Johan and Alon Clausen (2024). *johanoelgaard/AssetPricing: Asset Prices and Financial Markets, University of Copenhagen, Fall 2024*. en. URL: <https://github.com/johanoelgaard/AssetPricing> (visited on 12/19/2024).
- PyTorch* (2024). URL: <https://pytorch.org/docs/stable/index.html> (visited on 11/24/2024).
- Sak, Haşim, Andrew Senior, and Françoise Beaufays (Feb. 2014). *Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition*. arXiv:1402.1128. DOI: 10.48550/arXiv.1402.1128. URL: <http://arxiv.org/abs/1402.1128> (visited on 11/28/2024).
- scikit-learn (2024). *Decomposing signals in components (matrix factorization problems)*. en. URL: <https://scikit-learn/stable/modules/decomposition.html> (visited on 12/13/2024).
- Taylor G. Smith (2021). *pmdarima: ARIMA estimators for Python — pmdarima 2.0.4 documentation*. URL: <http://alkaline-ml.com/pmdarima/> (visited on 12/19/2024).
- Tibshirani, Robert (Jan. 1996). “Regression Shrinkage and Selection Via the Lasso”. en. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1, pp. 267–288. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/j.2517-6161.1996.tb02080.x. URL: <https://academic.oup.com/jrsssb/article/58/1/267/7027929> (visited on 11/24/2024).
- Trading Economics (2024). *Brent crude oil - Price - Chart - Historical Data - News*. URL: <https://tradingeconomics.com/commodity/brent-crude-oil> (visited on 12/12/2024).

- Trebbien, Julius et al. (Oct. 2023). *Probabilistic Forecasting of Day-Ahead Electricity Prices and their Volatility with LSTMs*. en. arXiv:2310.03339 [cs]. doi: 10.48550/arXiv.2310.03339. URL: <http://arxiv.org/abs/2310.03339> (visited on 12/16/2024).
- Yahoo Finance (2024a). *Brent Crude Oil Last Day Financ (BZ=F) Stock Price, News, Quote & History*. en-US. URL: <https://finance.yahoo.com/quote/BZ=F/> (visited on 12/12/2024).
- (2024b). *Dutch TTF Natural Gas Calendar (TTF=F) Stock Price, News, Quote & History*. en-US. URL: <https://finance.yahoo.com/quote/TTF=F/> (visited on 12/12/2024).

# A Methods

## A.1 Principal Component Analysis

The principal component analysis used in this paper is derived using the singular value decomposition as specified in Jolliffe (2002). To do so, we start by centring the data around its mean.

$$\bar{X}_{ij} = X_{ij} - \frac{1}{N} \sum_{i=1}^N X_{ij} \quad (\text{A.1})$$

We then use singular value decomposition (SVD) to solve the matrix factorisation problem and obtain the singular values of the centred data.

$$\bar{X} = USV' \quad (\text{A.2})$$

Where  $U$  and  $V$  are the orthogonal eigenvectors in the rows and columns space, respectively, while  $S$  is a diagonal matrix with the diagonal elements being the singular values,  $\sigma$ . We can then utilise the relationship between SVD and the covariance matrix that  $\Sigma = V'S^2V$ , where  $S^2$  is the diagonal matrix with the eigenvalues of the covariance matrix.

We then find the explained variance for component  $k$  as:

$$\text{Explained variance ration for } k = \frac{\sigma_k^2}{\sum_{i=1}^N \sigma_i^2} \quad (\text{A.3})$$

When solving for our principal components, we accumulate the explained variances, solving for the lowest number of components,  $m$ , reaching the minimum explained variance threshold,  $\eta$ :

$$\arg \min_m \sum_k = 1^m \frac{\sigma_k^2}{\sum_{i=1}^N \sigma_i^2} \geq \eta \quad (\text{A.4})$$

When the number of components,  $m$ , is found, we construct the projection matrix using the first  $m$  columns in  $V$ ,  $V_m$ , to reduce the dimensionality of the data to  $m$  by:

$$\tilde{X} = \bar{X}V_m \quad (\text{A.5})$$

Where  $\tilde{X}$  is the  $N \times m$  matrix with the  $m$  principal components.

## A.2 Assessing predictive performance

For the assessment of our models' predictive performance, we evaluate them based on two metrics: root mean square error and mean absolute error evaluated on an hourly basis as

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} \quad (\text{A.6})$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t| \quad (\text{A.7})$$

### A.3 Diebold Mariano Test

To compare the performance and assess if the models outperform the naïve forecast, we use the DM-test proposed by Diebold and Mariano (1995). It is performed by first calculating the loss differential as:

$$d_t = \mathcal{L}_X(e_{1t}) - \mathcal{L}_X(e_{2t}) \quad (\text{A.8})$$

Where  $e_{1t} = y_t - \hat{y}_{1t}$  and  $e_{2t} = y_t - \hat{y}_{2t}$  and  $\mathcal{L}_X$  is either the MSE or MAE loss functions. The hypothesis of equal performance corresponds to  $H_0 : \mathbb{E}(d_t) = 0$  and is tested against the two-sided alternative of  $H_A : \mathbb{E}(d_t) \neq 0$ . The DM-statistic is calculated as the mean over a weighted sum of the sample autocovariances.

$$DM = \frac{\frac{1}{N} \sum_{t=1}^N d_t}{\sqrt{\frac{\hat{\gamma}_d(0) + 2 \sum_{k=1}^{h-1} \hat{\gamma}(k)}{N}}} \quad (\text{A.9})$$

Where

$$\hat{\gamma}_d(k) = \frac{1}{N} \sum_{t=k+1}^N (d_t - \frac{1}{N} \sum_{t=1}^N d_t)(d_{t-k} - \frac{1}{N} \sum_{t=1}^N d_t) \quad (\text{A.10})$$

The DM-statistic asymptotically follows a Student's  $t$ -distribution with  $N - 1$  degrees of freedom under the null from which the p-values are computed.

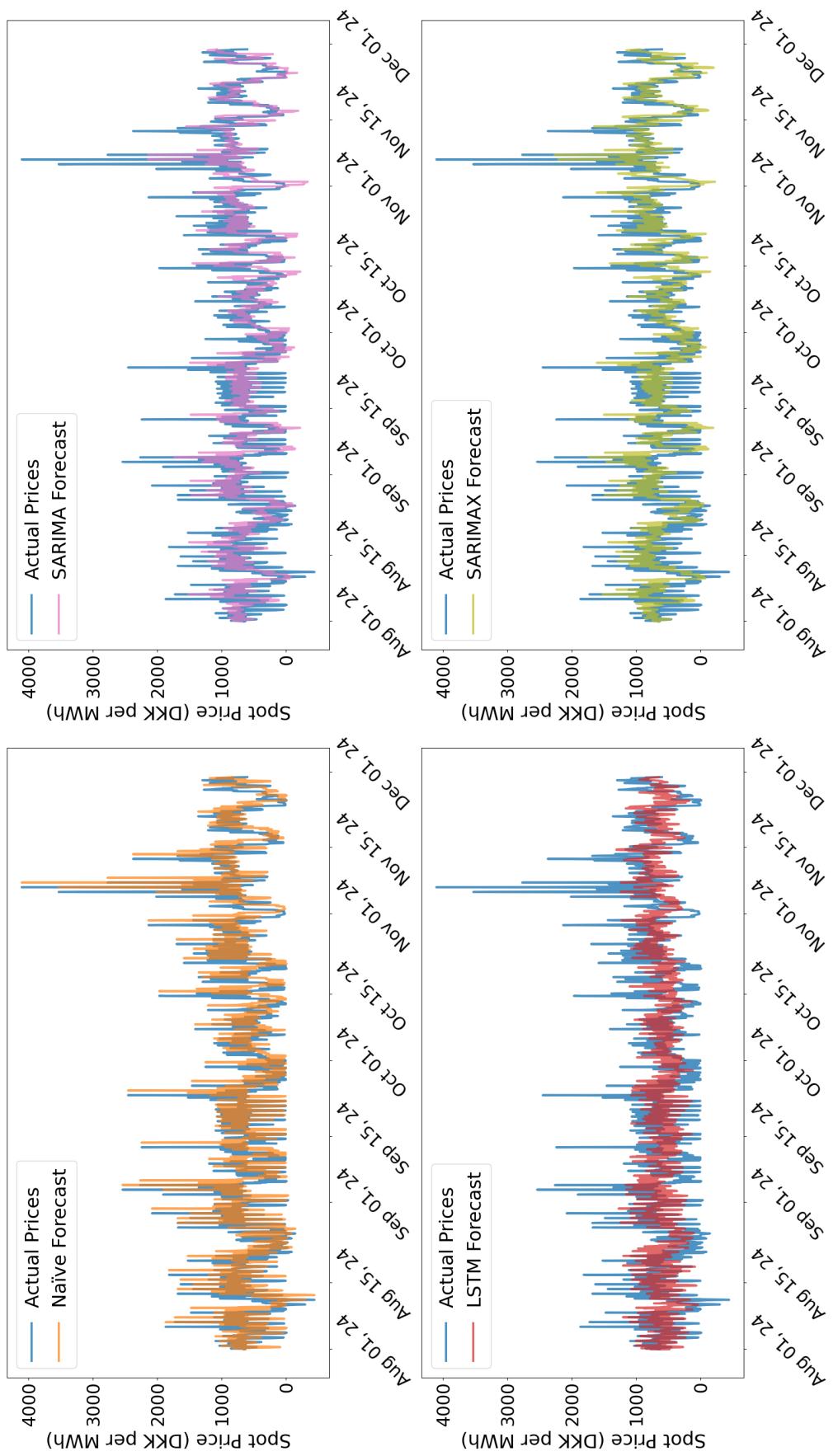
## B Estimation

Table B.1: Searched parameter values

Parameter	Values searched
$\lambda_{L1}$	$10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$
$\lambda_{L2}$	$10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$
hidden dimensions	128, 256, 512
hidden layers	2, 3
batch size	256
learning rate	$10^{-3}$

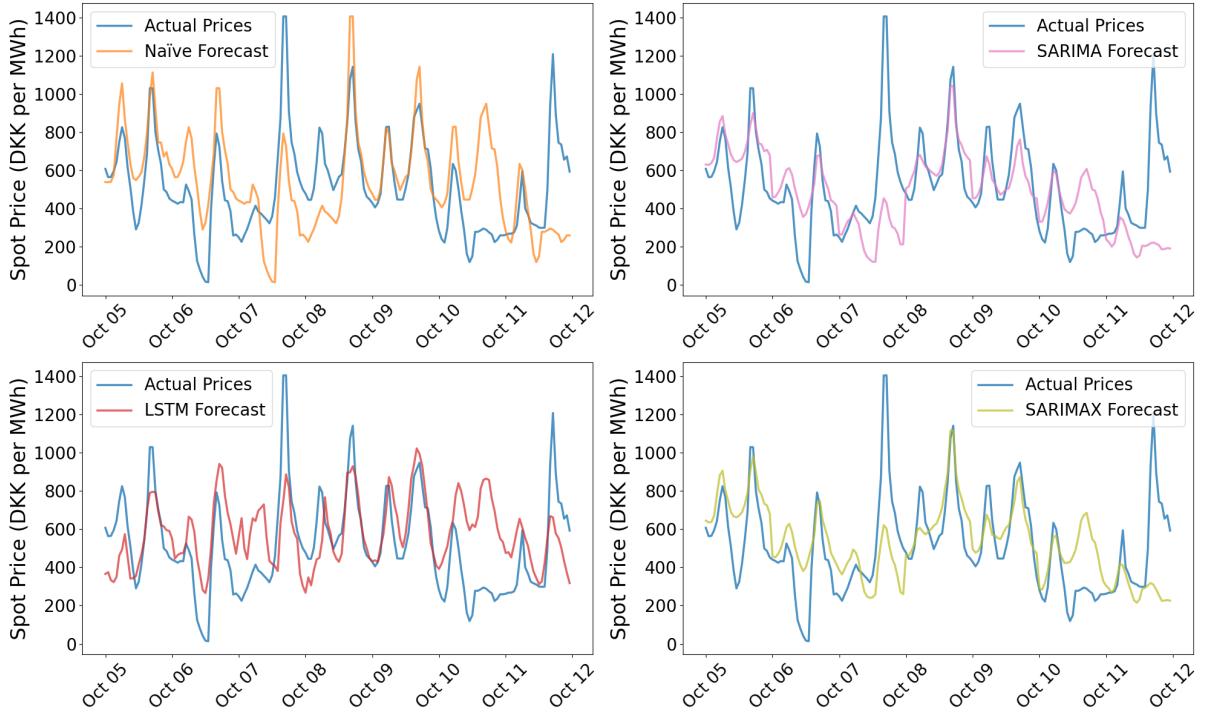
NOTE: Batch size and learning rates have been set to 256 and  $10^{-3}$ , respectively. We have tested batch sizes of 64 and 128 and learning rates of  $10^{-1}, 10^{-2}$ , and  $10^{-4}$  for a subset of the models, generally finding that the model performs much worse and have chosen not to pursue more rigorous testing. Further, we train the model for up to 100 epochs; however, if no improvement is seen in the validation data for ten epochs, we stop the tuning earlier. Generally, our LSTM NN stops before 50 epochs have passed

Figure B.1: figure  
Full predicted series for the naïve, SARIMA, SARIMAX and LSTM model



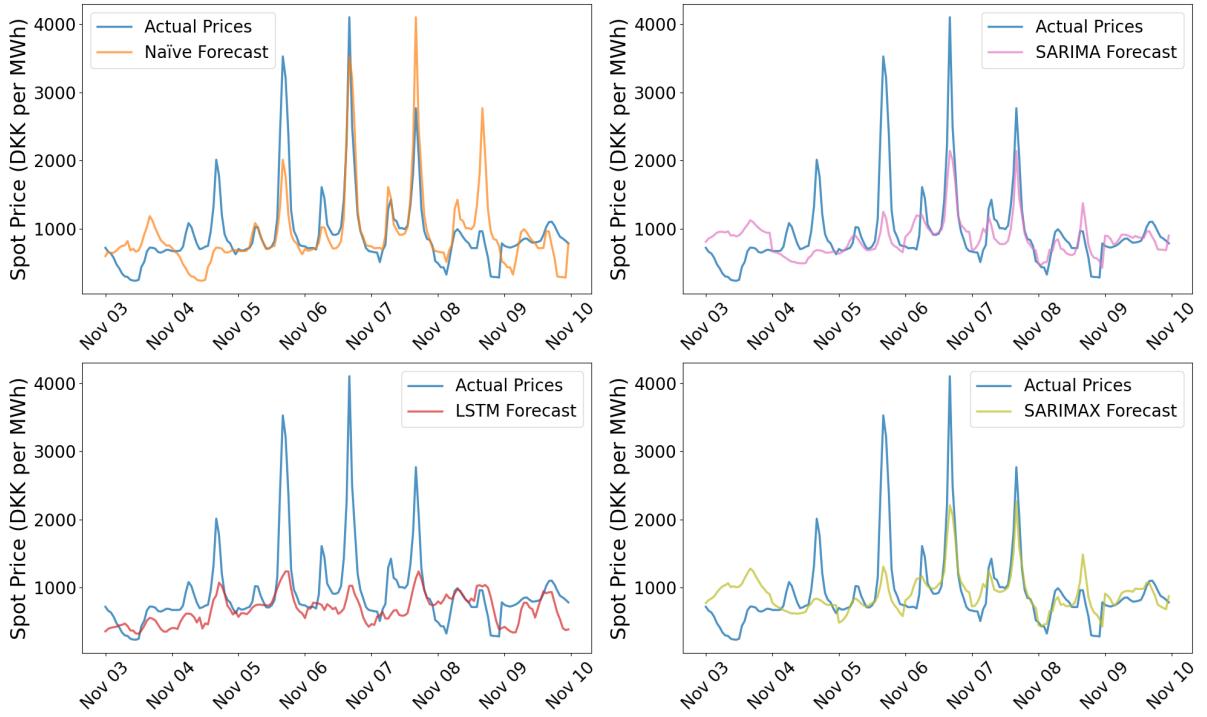
NOTE: Forecast values of the Naïve (top right), SARIMA (top left), SARIMAX (bottom left) and LSTM (bottom right)

Figure B.2: Forecasts for October 5<sup>th</sup> 2024 until October 12<sup>th</sup> 2024



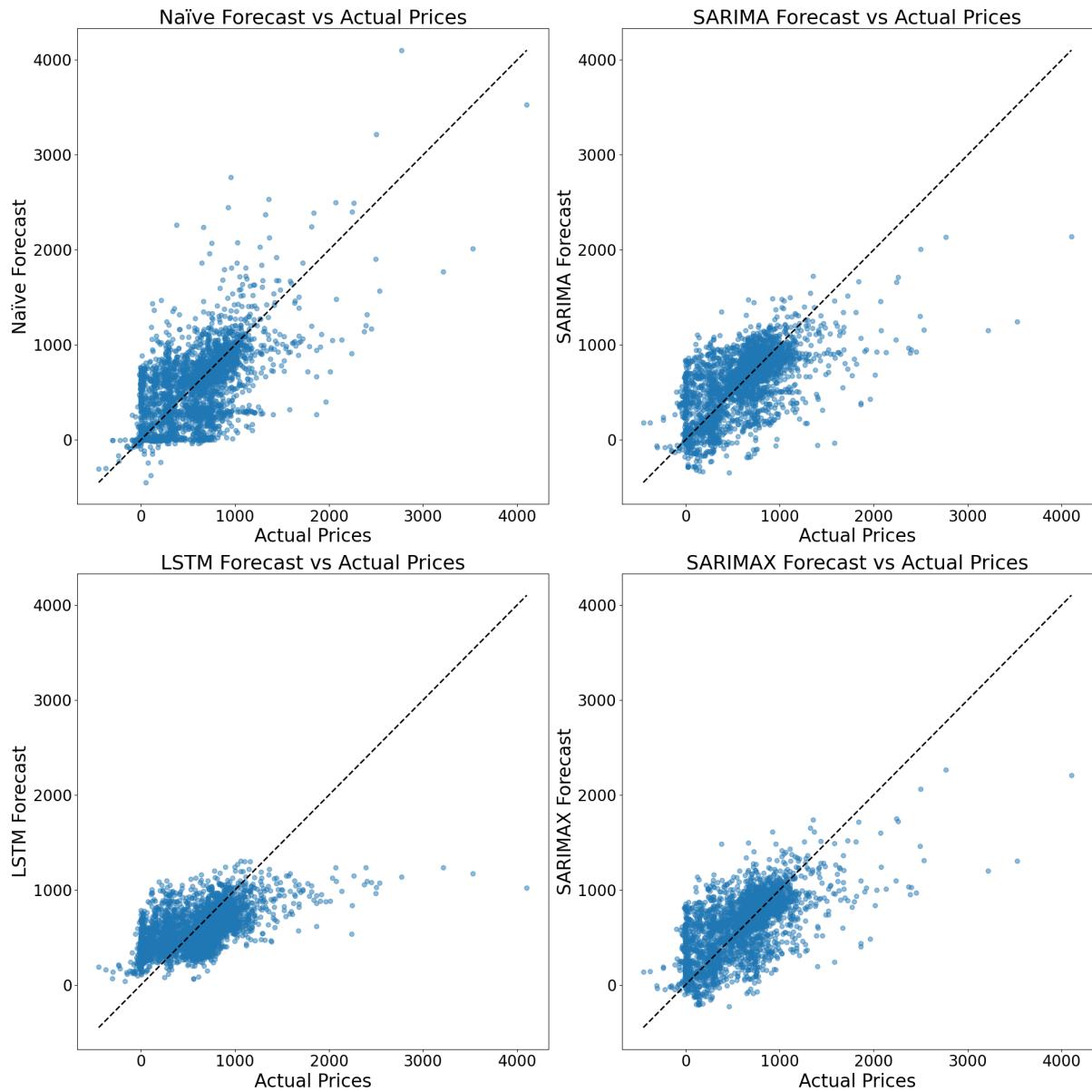
NOTE: Forecast for a period with relatively low volatility where the LSTM model slightly outperforms the others

Figure B.3: Forecasts for November 3<sup>rd</sup> 2024 until November 10<sup>th</sup> 2024



NOTE: Forecast for a period with high volatility. The naïve model seems to perform the best together with the SARIMA and SARIMAX models

Figure B.4: Scatter plot of forecast and actual values for November 3<sup>rd</sup> 2024 until November 10<sup>th</sup> 2024



NOTE: Scatter plot of predicted and actual values. We see the LSTM model keeping most predictions below DKK 1000 MWh, whilst the SARIMA and SARIMAX models can predict higher