Master Thesis Financial Economics

The changing dynamics of the Dutch Day-Ahead market, a modeling comparison with Covid-19 as a crisis-period

Using (S)AR(I)(MA)(X), ARIMA-GARCH and regime switching models with temperature and windvelocities as exogenous regressors

---

Abstract

In this study, the changing dynamics in the electricity market are modeled. The aim of the study is to compare the Covid-19 as a crisis period with the dynamics of the market in the 'normal' period during 2014 until 2020. The study compares in-sample performance accuracy using ARIMA, SARIMA, ARIMAX, ARIMA-GARCH and Markov-Switching-GARCH type models in both intraday and daily data-frequency. As an extension, are the extreme weather influences tested on the Dutch Day-Ahead market prices.

Author:	Daniel Delodder
Student ID:	545268DD
Supervisor:	dr. R. Quaedvlieg
Second Assessor	dr. R. Barahona
Date Final Version:	March 10th 2021

Keywords: Financial Econometrics, Model Construction and Estimation, Time-Series Models, Large Data Sets: Modeling and Analysis, Energy Forecasting

JEL-codes: C58, C51, C32, C55, Q47

Acknowledgement

I want to use this opportunity, to express my gratitude towards my supervisor. Rogier Quaedvlieg supported my wide range of ideas and helped narrowing them down effectively, especially on the more econometric side of things. Furthermore, did Ronald Huisman motivate me to write this thesis by his seminar 'Energy Finance', and did his book 'An Introduction to Models for the Energy Markets: The Thinking behind Econometric Techniques and Their Application' serve as a useful starting point for this thesis. Writing this thesis allowed me to take a deep dive in the energy market while learning many new theoretical concepts in econometrics, and practical skills in the program R. I truly enjoyed learning this combination of theory and practice, which I am sure of will help me in the future.

Contents

1. Introduction

The Dutch power markets were only fully denationalized in 2004. This rather young market still shows many irregularities and is highly subjective to change. One of the power market's characteristics is its rather volatile behavior. Where the WTI oil futures market showed its first negative price on the 20^{th} of April 2020, electricity markets such as the imbalance market show such negative prices on a near daily basis. The supply side of the electricity market is in transition resulting from the introduction of renewably generated power, one of the most significant contributors towards lower carbon emissions. Renewable power generation is more volatile in comparison to its current alternatives such as coal, nuclear and gas-generated power (Figure 5) which may cause the electricity market to become more volatile once the share of renewables grows. Furthermore, are costs of renewably generated sources such as wind and solar in the fixed part, while marginal costs of these renewable sources are considerably lower than their (fossil) alternatives (Figure 6). These two components make that, given the importance of growth of these renewable sources in the energy mix, the electricity market has been changing rapidly and will likely do so even more in the near future, with the risk of increasing volatility on the electricity grid.

Price volatility is a crucial feature for understanding spot price behavior of electricity, an important variable when pricing options and in macroeconomic models, and in particular when doing financial valuation of electricity derivatives. The relevance of accurately modeling electricity prices and energy prices in general has come to show a few times throughout the development of this study. During the Covid-19 pandemic demand for electricity and oil changed significantly, resulting in lower prices and increased volatility. Furthermore, did extreme temperatures cause friction in power markets. During the heatwave in California (August 2020), the state's power market struggled causing power blackouts. During the extremely cold weather in Texas in February 2021, it's power supply market proved unreliable causing the wholesale electricity price (operated by ERCOT) to spike by more than 10.000% on February 13^{th}, causing 4.5 million Texan households to be cut off from power. Accurate forecasting is named as part of the solution for prevention of such events.

When prices can be forecasted accurately, many practical applications arise. Energy companies controlling assets can improve decision making, option pricing can be performed more accurately, profitable trading schemes could be created and financial institutions can provide better informed financing.

As extreme weather circumstances will likely occur more often in the near future (Lubchenko and Karl, 2012), and power supply markets are expected to continue to develop towards more volatile renewable sources, away from stable fossil fuels (Figure 7). This validates the relevance of accurate modeling and forecasting.

In this we paper we want to examine which model-types model the Dutch Day-Ahead market best, with its changing dynamics, and how a crisis-period resulting from the Covid-19 period affects modeling performance, in order to obtain the high forecasting performance. The Covid-19 crisis period ('Post-C19 period') will be compared to the period 2014-2019 in both an intraday data frequency and in a daily frequency. In addition, do we want to examine these changing dynamics, add exogenous regressors into models and do we test the effect extreme windspeeds and extreme temperatures onto prices and whether these effects are increasing over the examined years. Therefore, this study follows the general research question:

"Which model-types model the dynamics of the Dutch Day-Ahead market best, and how does the Pre-C19 period compare to the Post-C19 period in terms of these criteria?"

In the next section important concepts will be explained, followed by a review of existing electricity price forecasting literature, a general description of the findings of previous research regarding the applied models is described in section 2.2, whereas in section 2.3 relevant methodologies in the energy market will be introduced. In section three, the methodology, the data will be described, and models that will be applied will be explained. Furthermore, will the model performance criteria be explained. In section 4, the results are described and the best performing models will be evaluated in section 4.3. Thereafter the conclusions, and implications end the study.

2. Literature Review

2.1 Conceptual Framework

Prior to assessing the literature, it is important to explain certain terminology, as interpretations in some terms can vary slightly in the literature.

2.1.1 The Energy Market

As a result of the non-storability of electricity prices, Day-Ahead prices show certain characteristics such as mean-reversion, seasonality, spikes and a complex time-varying volatility structure, Huisman et al. (2007) state. They lay focus on the characteristics of intraday electricity prices in Day-Ahead markets. The authors state that intraday electricity prices do not behave as a time-series process, but rather should be treated as a panel using 24 cross-sectional hours varying from day to day. They introduce a panel model for intraday electricity prices in Day-Ahead markets and show that the prices have specific mean-reversion and that they oscillate around an intraday specific mean price level. Moreover, do they find a block-structured cross-sectional correlation pattern in between hours to be present. Knittel and Roberts (2005) apply some models taken from traditional finance and fit these onto Californian electricity prices. They find the forecasting performance of traditional financial models is poor and can be significantly improved addressing the unique characteristics of the electricity price market. Think of non-storability, seasonality and peak-prices.

Weron (2014) provided a review of the literature of electricity price forecasting. The author finds basically five main approaches for electricity price forecasting, namely multi-agent models, fundamental models, reduced form models, statistical models and computational-intelligence models (Figure 3). The scope for this paper will lay on univariate statistical models with exogenous regressors.

Participants on the electricity market basically trade on a few main markets, the Day-Ahead market being the largest wholesale market in terms of volume. The Day-Ahead and imbalance market will shortly be explained, as their characteristics are important to note for this study.

2.1.3 The Day-Ahead Market

In literature the Day-Ahead market is often referred to as the ‘spot’ market. Compared to other commodity or financial markets this is unusual as in the Day-Ahead market does not allow for continuous trading. Agents in this market submit their bids and offers per hour, for electricity delivery the day after. This is important to notice when modeling and forecasting this ‘spot’ price. The market organizer (often the *TSO*, Transmission System Operator) matches supply and demand to determine a market clearing price every hour. Agents trade on this market clearing price regardless of their bid. Bids are placed and the clearing price is determined on day $t-1$, while physical delivery takes place on day t. Negative price bids are allowed in most markets (with exception to the Portuguese & Spanish markets), which occurs when supply peaks from renewable sources (often due to wind). The market clearing price is the price a bidder will actually pay for delivery of electricity, regardless of his own bid.

2.1.4 The Imbalance Market

In the imbalance market the time-horizons are much shorter, often per hour or half-hour and in the Dutch market even per quarter. The TSO operates the balancing market. This market is mainly used to price deviations in supply and demand from Day-Ahead or long-term contracts. The TSO uses this technical market to ensure the balance for Day-Ahead or longer-term markets, to match supply when deviations occur. This makes that the imbalance market has different characteristics than the Day-Ahead market. In recent years, a slight shift from the Day-Ahead to the imbalance market is visible in terms of volumes, partly because it allows for continuous trading. The imbalance market has more extreme peaks, have a lower price level, higher volatility and more extreme spikes[1]. Since this market acts as a balancing market, and given the more extreme characteristics of this market, forecasting and modeling has not been performed on a large scale in the literature as it is also considered more complex. There are some exceptions. Ma et al. (2004) develop neural network models for forecasting real-time Locational Marginal Prices (LMP) before and after the Day-Ahead market clearing. Olsson and Söder (2008) model imbalance prices using seasonal ARIMA and discrete

[1] Mean (2010-2020) Imbalance Market = €44,79, standard deviation. = €55,64
Mean (2010-2020) Day-Ahead Market = €53,01, standard deviation = €12,14

Markov processes and find the combination model appropriate. Accuracy is in general higher for Day-Ahead forecasts than for imbalance or other real-time prices, Aggerwal et al. (2009) conclude in a reviewing paper.

Terminology in the literature is flued and differs among regions, authors and years. The term 'spot' can was in the past mostly used to refer to the Day-Ahead market. In Europe convention is that to refer to the Day-Ahead as the spot, but in the US the spot price refers to the intraday (imbalance) market, while the Day-Ahead price is referred to as the forward price. In terms of terminology in this paper we will refer in line with US terminology to the Day-Ahead price as the forward price, and to the imbalance price as the spot price.

2.2 Literature

2.2.1 Forecasting and modeling - general

Weron (2006) provides an overview of modeling approaches, focused on practical applications of statistical methods for Day-Ahead forecasting which are ARMA-type, ARMAX and GARCH-types and moves on to quantitative stochastic models for derivatives pricing, namely jump-diffusion and Markov regime-switching. Zareipoor (2008) reviews linear time-series models (ARIMA, ARX, ARMAX) and nonlinear models (regression splines, neural networks) and uses these to forecast intraday power prices in the Ontario market. This is the clearest classification of models, where for example Haghi and Tafreshi (2007) classify ARIMA, ARIMA-Wavelet, ARX and ARMAX models as 'stationary', and neural networks, regime-switching, GARCH, jump-diffusion and mean-reversion type models as 'non-stationary' which is confusing since some models (such as mean-reversion) can in fact be stationary. In this thesis we will lay focus on reduced-form models and statistical models as characterized in Figure 3. Amjady and Hemmati (2006) acknowledge a limitation of time-series techniques, which they claim only to be successful in "areas where the frequency of the data is low, such as weekly pattern". Garcia-Martos and Conejo (2013) review short- and long-term electricity price forecasting and consider ARIMA and seasonal ARIMA calibrated for intraday daily-ahead price forecasts, and vector ARIMA (VAR in essence) and factor models for medium-term horizons. Given these contradicting views, assessing both intraday and daily frequency proves for interesting research.

Wolak (1997) studies formation of intraday prices. He finds that prices in the United Kingdom are more difficult to forecast compared to Nord Pool prices. This might however be due to the large share hydropower has in the energy mix in the Nord Pool, and electricity can be considered a storable commodity. Considering The Netherlands' energy mix, we find hydropower to represent a low share in the Dutch energy mix (Figure 5) and as having low potential. The low share and potential of hydropower in the energy is partially explained by the flat and often sub-sea level landscape and subsequent limits for hydro powered resources.

In terms of modeling and forecasting horizons one can distinct between short-term, medium-term and long-term. Short-term forecast horizons are most important for active traders and to market operators. Medium-term is considered one to a few days ahead. Medium-term horizons are used most for option pricing and risk management. Long-term forecasts are considered weeks, months and even years ahead. This type of horizon is mostly used in financing and planning of operating power plants. As can be expected, the significance of these long-term horizons decreases. In this paper short and medium-term will therefore be in scope. Higher frequency data can be better used for medium-term forecasts, high-frequency data can be better used in intraday forecasts.

2.2.2 Fundamental Models

Fundamental or structural models capture economic relationships and characteristics in electricity trading and production. Influences of weather, load or other fundamental drivers are modelled and predicted independently, via statistical, reduced-form or computational intelligence techniques, so often hybrid solutions of the model categories as presented in Figure 3. Fundamental models can be sub-categorized into parameter-rich models, and parsimonious structural models of supply and demand. The parameter rich models are mostly used in hydro-dominant power markets (Nordics and Switzerland) to explain certain parameter influences onto prices such as snow and temperature. Parsimonious structural models are constructed from supply and demand supply curves analyses. Kanamura and Ohasi (2007) construct a supply curve intercepting a mean-reverting demand curve and show a rather ‘spikey’ electricity price process which fits the PJM prices (Northeastern-US’s power market) better than a jump-diffusion model. The authors later improve their model by adding transition probabilities. Boogert and Dupont (2008) apply a supply-demand framework to the Dutch Day-Ahead market and forecast 24 hours ahead. They find that adding the reserve margin enhances performance.

Fundamental models can be based on stochastic processes such as demand and generation capacity. The spot price process can capture spikes and complex dependence on underlying factors. Data availability is an issue. Many other variables besides prices are only available in longer timeframes, which makes this type of modeling most suited for medium term forecasting and are mostly used for derivatives pricing. Secondly is the stochastic behavior of the fundamental factors difficult to model.
The focus in this study will not be on fundamental models, but is described for general understanding of the landscape of the forecasting and modeling in the electricity market.

2.2.3 Statistical Models

Following the classification of Weron (2014), exponential smoothing, regression models, AR-type and ARCH & GARCH-type models are considered statistical (econometric) modeling approaches. When discussing reduced form models it was concluded that these types are outstanding models to use for derivatives valuation and risk management. But its simplicity limits the forecasting performance. Statistical (econometric and technical analysis) methods forecast current prices by a mathematical combination of the previous prices and/or previous exogenous factor values. (load, consumption, weather or production for example).

This (S)AR(IMA) type of modeling is not most used in the literature but can perform well in short-term forecasts. Typically forecasts become more accurate when each hour is modelled separately and when simple probabilistic processes for jumps are included, Weron (2014) states.
However, as discussed previously the electricity market is shows strong characteristics and depends on various exogenous variables such as weather conditions and capacity. In order to capture these exogenous factors and their influence when modeling, while building on the (S)AR(IMA) platform as derived, (S)AR(IMA)(X) models come into play (where the added X refers to exogenous variables).

Still one of the most widely used types of modeling, also in electricity price forecasting, are regression models. In the literature they are often combined into a hybrid model with more sophisticated methods. Several papers, such as the Karaktsani and Bunn (2008) paper conclude that models that incorporate market fundamentals and time-varying coefficients outperform in terms of forecasting performance.
Johnsson et al. (2013) use a two-step methodology. First, they examine the impact of system load and wind power generation. These (nonlinear and nonstationary) explanatory variables their influence are accommodated in a nonparametric and Time-Varying Regression model. In the second step an AR model and exponential smoothing are applied to account for residual autocorrelation and seasonal dynamics.

Wang and Wu (2012) forecast energy market volatility using univariate and multivariate GARCH models. They investigate the volatility dynamics of energy prices focusing on crude oil and find that multivariate models display better performance than univariate models. Their evidence shows that univariate models allowing for asymmetrical effects display greatest accuracy. Sadorsky (2006) finds the threshold GARCH to fit well for heating oil and natural gas volatilities (using daily futures of the petroleum market) and GARCH(1,1) to fit well for crude oil and unleaded gasoline volatilities. Wang and Wu (2012) test if Sadorsky's findings hold on spot markets since they are less liquid, show less market noise and are subject to higher transaction costs, and find that all these factors may influence energy price volatility persistence and thus influence the modeling performance. When applying Superior Predictive Ability (SPA) test, Wei et al. (2010) argue that none of the GARCH models can outperform all of the others. Fong and See (2002) state that a common finding is that GARCH models tend to impute a high degree of persistence to the conditional volatility, implying shocks in the distant past continue to have non-trivial effects on current prices. The GARCH method provides us with an indication as to whether the volatility is time-varying and give weight to the importance of yesterdays' volatility and can improve upon the limitations of ARIMA-type models as they are limited by the assumptions of linearity and homoscedastic error variance (Lama et al. 2015). A limitation of GARCH models is that the parameters need to be estimated which may be a rather time intense method and subject.

Empirical evidence for GARCH behavior in electricity is rather limited compared to price behavior for other forms of energy (Duffie and Gray, 1995). Haas et al. (2004) include regime changes in their Markov Regime Switching GARCH (MRS-GARCH) models to allow for nonlinearities associated with time-varying volatility applied in the FOREX market, overcoming the mentioned estimation difficulties. Their results show this model tends to have greater accuracy than more conventional methods. A limitation of pure GARCH modeling is that GARCH-type models assume the series has no autocorrelation, no drift and no seasonal effects. This can all be incorporated the conditional mean component, by using (S)AR(I)MA(X) models for the mean and continuing modeling with the residuals as the conditional volatility. Combined ARIMA-GARCH models can improve forecasts as it adds time-varying volatility.

2.2.4 Reduced-form Models – (Markov) Regime switching & Jump-diffusion type

These quantitative stochastic type model approaches characterize the statistical properties of electricity prices over time, targeting derivatives evaluation and risk management as implementation objectives. Reduced-form models are mostly used to replicate the main characteristics of electricity price-dynamics rather than obtaining an accurate price forecast. These models are used in risk management and derivatives pricing.
Weron (2014) considers reduced-form models accurate forecasting methods, though recover the main characteristics of electricity prices often. The models provide a simplified but reasonably realistic picture of the price dynamics often using a daily timeframe. With regards to volatility or price spikes forecasts, Weron (2014) argues reduced-form models perform reasonably well.

Huisman and Mahieu (2003) study the Dutch APX (Day-Ahead) market and argue that a stochastic jump process with mean-reversion might lead to an erroneous specification of the true mean-reversion process. They show the existence of a 'normal mean reverting process' that is not directly associated with jumps. In the regime jump model, they set-up three different regimes. They identify a normal regime that can contain a mean-reversion component. The first regime models a price jump and a second regime models the way the process falls back to the normal process. Markov transition matrices specify the probabilities that prices move in-between two regimes. They argue the advantage compared to a stochastic jump model is that the short-lived characteristic of power prices can modelled.

Mount et al. (2006) show a stochastic regime switching model to represent the volatile behavior of wholesale electricity prices associated with price spikes effectively. The authors introduce the flexibility of the model's structure and amend Ethier and Mount (1999) and Hamilton (1989) their findings that the Markov regime-switching model can be used to capture the behavior of electricity spot prices, by that the Markov regime-switching between a high-price regime and a low-price regime allows for stochastic jumps of prices. Each regime is a mean reverting AR(1) process. The authors find that the estimated parameters have distinct seasonal patterns and conclude that a model with time-varying coefficients would be more appropriate.

Furthermore, do they show that the estimated switching probability from the low to the high regime predicts price spikes well when the reserve margin is measured accurately.

The volatility term is usually set to a constant for simplicity, despite that electricity prices are heteroskedastic[2].

Since simple mean-reversion models do not allow for price spikes to occur and remain for a longer period, a model allowing for jumps by describing two different regimes might fit the data better and be used for forecasting. Fong and See (2002) combine GARCH and Markov-switching models to examine the temporal behavior of volatility of daily crude oil futures returns using a generalized regime switching model that allows for abrupt changes in mean and variance, GARCH dynamics, basis-driven time-varying transition probabilities and conditional leptokurtosis. They show regime shifts are clearly present in the data and dominate the GARCH effects. Their options framework relies on demand or supply shocks to explain the increase in the volatility of spot oil prices while the theory of storage places more emphasis on low inventories and volatile demand shocks as the main factors behind big swings in oil prices.

2.3 Methodology Overview

2.3.1 Introduction

In the methodology overview we shortly want to provide an overview of the existing models found in the literature, that model certain characteristics of the power market best. This serves as an intermediate step between the literature review and the models as they will be used in this study.

Introduction models

Huisman (2009) provides an introduction to modeling in the energy markets. This book provides an overview of methods used in the energy market, and how different characteristics can be incorporated when modeling. It provides background and builds up toward the models that will eventually be used for this study.

Energy prices are known to be mean-reverting, which implies that the price of a commodity hovers around some mean-price level (which may vary over time). Swartz and Smith (2000) explain why mean reversion exists, as when prices are low, supply will decrease as high-cost producers will exit leading to a downward pressure on prices (and vice versa), balancing the changes. Knittel and Roberts (2005) show that mean-reversion is not stable over the day, super-peak hours (18 to 22) exhibit less mean-reversion. They state this can be explained by the higher demand for power in these hours resulting in less reserve production capacity and therefore an increased probability of shortages and spikes. During these hours, prices are less predictable.

2 Heteroskedasticity implies that the residuals' variance is unequal over a range of measured values.

Let S_t be the natural logarithm of the Day-Ahead price of 1 MW of electricity delivered in day $t + 1$ such that t denotes the date of quotation instead of the delivery date. Assume that the spot price is the sum of two components: a deterministic component or forecastable component f_t and a stochastic component X_t. The deterministic component contains the forecastable dynamics of prices such as seasonal behavior or mean-price levels. The stochastic component models the stochastic behavior of the price net of the deterministic component. The spot price therefore equals:

$$S_t = f_t + X_t$$

The deterministic component f_t is assumed to equal some equilibrium mean-price level μ, consistent with the motivation of Swartz and Smith (2000):

$$f_t = \mu$$

The stochastic price factor X_t makes the price hover around the deterministic price f_t. Stochastic price factor is:

$$\partial x_t = -\propto X_{t-1} + \sigma\varepsilon_t$$

Where ∂x_t is the difference between X_t and X_{t-1}. The volatility of the stochastic price changes equals σ and is assumed to be constant here, although we are going to change this by making it time varying later. Mean reversion is captured by the $\propto$ component should be between 0 and 1. As $\propto$ approaches 1, stochastic prices will fall back to 0 more quickly and vice versa. Parameters $\propto$, μ and σ can be estimated using maximum likelihood.

ARIMA-type

The AutoRegressive Moving Average type model (ARMA (p, q)) is a standard time series model which accounts for time correlations. Let X_t be current price in the autoregressive part, linearly explained in terms of its p past values. Let q be previous values of the noise which stands for the moving average part:

$$\phi(B)X_t = \theta(B)\varepsilon_t$$

B stands for the backward shift operator. This implies that $B^h X_t = X_{t-h}$, so $\phi(B)$ stands for the function $\phi B = 1 - \phi_1(B) - \cdots - 1 - \phi_p B^p$ where similarly $\theta(B)$ stands for $\theta B = 1 + \theta_1 B + \cdots + \theta_q B^p$ where ϕ_1,…, ϕ_p and θ_1,…, θ_q are coefficients of AR and moving average polynomials (depending on software package). The noise variable ε_t is denoted as $WN(0, \sigma^2)$ implying zero mean and finite variance. For $q = 0$

this is an AR(p) model, and for $p = 0$ it is a moving average model MA(q). ARMA modelling assumes (weak) stationarity. If data is not stationary, the autoregressive integrated moving average (ARIMA) modeling approach offers a solution. The ARIMA type model has three parameter types, namely the above derived AR, MA and adds the number of differencing passes at lag-one (d). This ARIMA type of modelling is denoted as ARIMA(p, d, q), and is formally written:

$$\phi(B)\,\Delta^{d}\,X_t = \theta(B)\varepsilon_t$$

Here $\Delta^{d}\,X_t \equiv (1 - B)x_t$ is the lag-1 differencing operator, which allows for more general lag-h differencing operator $\Delta_h\,X_t \equiv (1 - B^h)x_t \equiv x_t - x_{t-h}$. Not always does differencing at lag-1 suffice to make the series stationary. Especially given the seasonality in electricity prices, longer lags can be of essence. In the literature this is known as seasonal ARIMA models (SARIMA).

2.3.2 Seasonality

Energy prices follow different price patterns throughout the different seasons due to weather differences. To incorporate seasonality in the model, we assume seasonality is a forecastable component and therefore included in the deterministic part f_t. Huisman and Mahieu (2003) incorporate a weekend seasonality factor, let this be w_t be a dummy variable that will be 1 if $t + 1$ is a weekend delivery day and 0 if not.

$$f_t = \mu + \beta w_t$$

Parameter β represents the amount by which spot prices differ from the mean price on weekend days or holidays. Huisman (2009) estimates β to be -0.48 for Dutch APX Day-Ahead electricity prices, and in general is expected to be negative.

Often are daily prices used in research. Huisman et al. (2002) propose an hourly model and explain that the Day-Ahead's hourly prices have to be considered a cross-sectional model of 24 individual prices, which are likely to be correlated. The authors also explain that when using hourly prices some significantly different characteristics of the energy market are described. When using daily averages of for example the Day-Ahead market, it serves well as a reference price for making market valuations and provide a base for option contracts, but average prices do not meet the microstructure of the Day-Ahead market itself. They propose the following panel model using hourly prices:

$$s_{h,t} = f_{h,t} + x_{h,t}$$

$$f_{h,t} = \sum_{i=1}^{24} \mu_i I_h^i + \sum_{d=1}^{6} \beta_d I_t^d$$

$$x_{h,t} = x_{h,,t-1} + \alpha_h\left(0 - x_{h,t-1}\right) + \varepsilon_{h,t}$$

$$\varepsilon_t \in [0, \Sigma]$$

Subscript h, t stands for an hourly (h) observation at delivery day (t). $f_{h,t}$ stands for the deterministic component as applied in Huisman et al. (2007). μ_i being the mean-price level for each hour $i = 1, \ldots ,24$. I_h^i is a dummy variable which will turn to 1 when i equals h and 0 otherwise. The second component β_d of the equation contains the daily deviations from the mean-price level, again as a dummy variable which turns to 1 during weekend days.

The $x_{h,t}$ component stands for the mean-reversion model delaying with the stochastic price behavior for hour h for succeeding days. In this part the prices are allowed to follow a time series specification over the days but incorporating hourly characteristics. The model allows for a speed of mean reversion α_h that is specific for hour h. The correlation within a day between the hourly stochastic prices is captured through Σ, which is a (24 x 24) correlation matrix in which cell (i, j) equals the correlation between the stochastic price of hours i and j. The error term $\varepsilon_{h,t}$ is therefore a (24 x 1) vector with hourly specific error $\varepsilon_{h,t}$ in line with h.

2.3.3 Volatility

Up until this point we considered the standard deviation of the error term to be constant. Real electricity prices do not confirm this and show variation in prices at a different rate. Incorporating time-varying volatility will likely increase the strength of results, as the random component in the decomposed data in Figure 9 & Figure 10 indicating volatility to follow a time-varying process.

The generalized autoregressive conditional heteroskedasticity (GARCH) type model as developed by Bollersev (1986) is a way of specifying time-varying volatility. The GARCH (1,1) is specified as:

$$\sigma_t^2 = \phi_0 + \phi_1 \epsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2$$

The GARCH(1,1) method puts the value of ϕ_1 onto the most recent observation. If ϕ_1 and ϕ_2 are zero, volatility is constant and equal to ϕ_0. The GARCH(1,1) method will by using the estimated parameters, detect the variability of volatility over time. We can add factors to the GARCH method, for instance a GARCH(2,2) method adds ϵ_{t-1}^2 , ϵ_{t-2}^2 and $\sigma_{t-1}^2, \sigma_{t-2}^2$.

Fong and See (2002) state that a common finding is that GARCH models tend to impute a high degree of persistence to the conditional volatility, implying shocks in the distant past continue to have non-trivial effects on current prices. The GARCH-method provides us with an indication as to whether the volatility is time-varying and gives weight to the importance of yesterdays' volatility. It is therefore able improve upon the limitations of ARIMA-type models, as they are limited by the assumptions of linearity and homoscedastic error variance (Lama et al, 2015). However, a limitation is that the parameters need to be estimated which may be a rather time intense method and subject.

2.3.4 Incorporating spikes

Mean reversion models are the basis for modeling the energy market. In reality the energy market might be more complex, as large spikes in prices may occur as a result of underlying changes in demand or supply. Models should capture these changes. The mean-reversion models did not account for the large spikes which the energy market is known for. The stochastic price factor ∂x_t can be extended, as Huisman (2009), reviewing previous work, describes:

$$\partial x_t = \propto (0 - x_{t-1}) + \varepsilon_t + J_t \partial N_t$$
$$\varepsilon_t \sim [0,1]$$

Where J_t describes the jump size as introduced by Askari and Krichene (2008) which is assumed to be normally distributed and N_t describes the jump Poisson process, which will equal the total number of jumps that have occurred between time periods t-1 and t. Now jumps in prices would be permanent. It is not specified what happens after a jump. The mean reversion parameter $\propto$ has a force that pulls prices back, but the mean-reversion parameter would jump up along with the price spike. Mean reversion parameter $\propto$ is not well specified. Geman and Roncoroni (2006) developed the jump-reversion model in order to solve these problems. They added a component to the mean-reversion specification:

$$\partial x_t = \propto (0 - x_{t-1}) + \varepsilon_t + h(t) J_t \partial N_t$$

Variable $h(t)$ models the jump; +1 for a jump up and -2 for a downward jump, depending on the current current price level. The J_t (jump size) variable is however is not assumed to be normally distributed, as the sign of the jump is determined by $h(t)$. They find convincing support for the model to fit the data well. This is not a straightforward application of maximum likelihood.

Hamilton (1989) introduced the regime-switching model to include spikes in models. He assumed that the modelled variable can occur in more than one state. In energy markets this is applicable, since sometimes there can be short-term friction in supply causing large spikes in prices. During these periods, prices may behave differently compared to the 'normal' state. Mount et al. (2006) and Huisman (2003) use this framework and assume that the day ahead market switches between two states: a normal, mean-reverting state and a non-normal state in which price spikes happen. Following Hamilton (1989), Huisman (2008) and Mount (2006), an applicable framework will be constructed. Starting from (3), 2 different periods are added in the equation, and the meaning of σ will be changed, such that:

$$\partial x_t = -\alpha X_{t-1} + \sigma_1 \varepsilon_{1,t}$$

Now the 1 refers to being in State 1. However, when in State 2, we assume prices do not follow a mean reversion process. The price may be high or low. The stochastic price may suddenly jump:

$$\partial x_t = \mu_2 + \sigma_2 \varepsilon_{2,t}$$

Now, the mean-reversion parameter $\propto$ is only present in State 1, during normal market conditions where no spikes happen. In order to now specify how the process switches from the normal mean revering state 1 to the state 2 where spikes occur. Let s_t be the determinant of the state at time t, $s_t = 1{,}2$. With p the probability is specified by which the market was in state i at time $t-1$ and moves to state j at time t: $p_{i,j} = \Pr\{S_t = j \mid S_{t-1} = i\}$. So $p_{11}$is the probability that the market stays in State 1 and p_{21} is the probability that a spike happens.

The parameters can now be estimated using maximum likelihood. To do this the likelihood of observing the data $l(dx_t|\Omega_{t-1};\Theta)$, where Θ stands for all variables to be estimated. Using time periods until one lag, the probability that the value of ∂x_t can be observed needs to be estimated while excluding it's most recent observation. There are now four states that can occur, namely state 1, moving from state 1 to 2, moving back and state 2. Hamilton (2005) lets $\xi_{1,t-1}$ be the probability that the market was actually in state 1 at time $t-1$:

$$\xi_{1,t-1} = \Pr\{S_{1,t-1} = 1 \mid \Omega_{t-1};\Theta\}$$

$$\eta_{1,t} = f(\partial x_t \mid S_{1,t-1} = 1 \mid \Omega_{t-1};\Theta)$$

Where $\eta_{1,t}$ is the probability that ∂x_t can be observed in state 1. The four different states are aggregated:

$$l(\partial x_t \mid \Omega_{t-1};\Theta) = \sum_{i=1}^{2}\sum_{j=1}^{2} \xi_{i,t-1} p_{ij} \eta_{j,t}$$

Each likelihood depends on the previous state's probabilities. Now only $\xi_{j,t}$needs to be calculated. Hamilton (2005) shows:

$$\xi_{j,t} = \frac{\sum_{i=1}^{2} \xi_{i,t-1} p_{ij} \eta_{j,t}}{l(\partial x_t \mid \Omega_{t-1}; \Theta)}$$

The equation that the total probability that ∂x_t is observed in state j at time t. The denominator is the total probability of observing ∂x_t in all states. It is the marginal probability of observing ∂x_t in state j at time t. The next step would be to estimate $\xi_{1,t}$ and $\xi_{2,t}$.

Huisman (2008) indicates that using more regimes than two might lead to estimation problems. Mount et al. (2006) propose a regime-switching model in which the probability of a spike occurrence is time-varying. The authors show that the probability of a spike occurrence in the PJM market depends on the reserve margin. This resonates well with today's negative oil price spikes. As reserves are lower, the chances of a spike is higher. They specify the probability of remaining in regime i $(= 1{,}2)$ as:

$$p_{ii,t} = c + \frac{d}{rm_t}$$

Here rm_t is an estimate for the difference between the installed electricity production capacity and consumer demand or load; the reserve margin. To implement the reserve margin, accurate data about the reserve margin (the difference between available capacity and capacity in use) is required. Huisman (2008) uses daily temperature as a proxy for the reserve margin, as shocks in power consumption are related to shocks in temperature and much more easily available. Huisman (2008) finds that temperature and deviations from expected temperature levels influence the probability of a spike.

Fong and See (2002) model crude oil spot price volatility using a regime-switching model and compare the estimates to a GARCH (1,1) model and find the regime-switching model to fit data significantly better. Mount et al. (2006) start from Hamilton's (1989) original model looked like:

$$y_t - \mu_{s_t} =_{\Phi} \left(y_{t-1} - \mu_{s_{t-1}}\right) + \varepsilon_t$$

The most important feature of the model Mount et al. (2006) propose is that the key parameters are functions of observed explanatory variables. The authors amend that the general specification for the model can be written as follows:

$$y_t \; is \; N(\mu_{1t}, \sigma_t^2) \; if \; S_t = 1$$

$$y_t \; is \; N(\mu_{2t}, \sigma_t^2) \; if \; S_t = 2$$

Here y_t is the logarithm of price; μ_{i_t} is the mean for regime $S_t = i$ and σ_i^2 is the variance for regime $S_t = i$. In regression form, the model for y_t is mean reverting and a function of a vector of additional regressors. The form of the mean corresponds to the following geometric log model:

$$y_t = \mu_{it+}\varepsilon_{it} = \propto_i + \phi_i y_{t-1} + \gamma_i x_t + \varepsilon_{it}$$

where $\propto_i$, ϕ_i and γ_i are unknown parameters for regime $S_t = i$; $x_t$is an explanatory variable (additional variables can be included); ε_{it} is an unobserved residual that is $N(0, \sigma_i^2)$. The implication of (2) is that the conditional mean of yt, E[y_t|xt, y_{t-1}, S_t =i], varies with time. They finally allow for the two states and their two transition probabilities:

$\Pr[S_t = 1 | S_{t-1} = 1] = P_{1t}$
$\Pr[S_t = 2 | S_{t-1} = 1] = 1 - P_{1t}$
$\Pr[S_t = 2 | S_{t-1} = 2] = P_{2t}$
$\Pr[S_t = 1 | S_{t-1} = 2] = 1 - P_{2t}$

The unconditional variance of y_t varies with time because the probabilities of being in one or the other regime and the conditional means for each regime varies over time.

$$Var[y_t|\phi_t] = E[y_t^2|\phi_t] - E[y_t|\phi_t]^2 = \rho_t(y_{1t}^2 + \sigma_1^2) + (1 - \rho_t)(\mu_{2t}^2 + \sigma_2^2) - (\rho_t\mu_{1t} + (1 - \rho_t)\mu_{2t})^2$$

Here $\phi_t = [y_1, y_2, \dots, y_{t-1}, x_1, x_2,, \dots x_t]$ represents the information available to make a one-step ahead forecast of y_t in both regimes. The conditional probability of being in state 1 is given as $\rho_t = \Pr[S_t = 1 | S_t = 1 | \phi_t]$.

They end up with their Model 1:

$$y_t = \propto_i + \phi_i y_{t-1} + \gamma_i R_t + \omega_i \log(L_t) + \varepsilon_{it} \qquad for \; i = 1,2$$

The transition probabilities:

$$P_{it} = \Pr[S_t = i \,| S_{t-1} = i] = \frac{\exp\left[c_i + d_i\left(\frac{1}{R_t}\right) + e_i\left(\frac{L_t}{10000}\right)\right]}{1 + \exp\,[c_i + d_i\left(\frac{1}{R_t}\right) + e_i(\frac{L_t}{10000}} \qquad for \; i = 1,2$$

Where y_t is the logarithm of price at time t, R_t is the reserve margin at time t, L_t is the load in MW at time t and $\varepsilon_{it} \sim N(0, \sigma_i^2)$ for $i = 1,2$. The load accounts for seasonality and weekend characteristics.

3. Methodology

3.1 Data

Data has been gathered through Bloomberg. The data describes the separate intraday Day-Ahead auctioned prices from the period 2010 until 2020/09 for weekdays (so excludes weekends). Data for temperature and windspeeds are gathered through the KNMI, the Dutch public meteorologic institute. Windspeeds are in average intraday windspeed in 0.1 m/s, and temperature is in 0.1 degrees Celsius. Intraday and averages of the intraday prices are used in the regressions, and weekends are deleted to form the data in same frequency. All data is from the De Bilt station, the mainly used weather station in the Netherlands due to its central geographical location and representation of average countrywide weather.

As is shown in the decomposition of the time-series in Figure 9 & Figure 10, there is strong seasonality and a trend visible in the dataset, as was to be expected from findings as reported in the Literature Review. For the purpose of this study, we chose not to extract the trend and seasonality but to model these characteristics as well as possible. As the purpose of this study is to find the best fitting models to the electricity market given these characteristics and in addition that model performance will be even more visible in the raw data and that differences in will be more strongly present. Better in-sample performance could no doubt have been achieved when modifying the data, but given our purposes of modeling these characteristics we chose to use the raw data.

3.2 Descriptive statistics

In Table 1 the descriptive statistics and Augmented Dickey-Fuller test values are depicted of both the intraday and daily samples for the prices, and for the exogenous regressors temperature and windvelocity are described. The data is decomposed in both hourly and daily data frequency in Figure 9 and Figure 10 respectively. The data variables are explained below.

Variable	Description	Source
P_t	Dutch Day-Ahead price in period t, 24 hourly prices on a day. In daily sample the average of 24 daily observations is computed. These prices will be referred to as 'intraday' and 'hourly' (Day-Ahead) prices.	Bloomberg
WV_t	Windvelocity hourly average windvelocity in 0.1 m/s in De Bilt. Windvelocities are reported in rounded numbers to tenths (a windvelocity of 26 would be rounded to 30).	KNMI

T_t	Temperature in 0.1 degrees Celsius in De Bilt	KNMI
$VarP24_t$	The variance is a variable which is computed as the average variance among the observation in period t, and the 23 prior observations.	Computed
Dummy peak	The dummy for peak hours is set equal to 1 during hours 8, 9, 10, 11, 12, 13, 18, 19, 20, 21. The remaining hours are considered off-peak hours and are set equal to 0. This division is based on findings presented in Figure 4.	
Dummy winter	The months that positively deviated from the average Day-Ahead price over the period 2014-2020 resulted in the following 'summer' months: May, June, July, August, September, October. The remaining are considered summer months.	

Period	Description
Pre-C19	The Pre-C19 period is the period from 2014 until 2019.
Post-C19	The Post-C19 period is the period starting March 23 until May 29. This coincides with the formal Dutch 'intelligent lockdown' as a result of the C19 pandemic.

Table 1: Descriptive statistics and ADF-test statistics

Period	**Time-frame**	**Variable**	**Obs**	**Mean**	**Stdev**	**Min**	**Max**	**Skew**	**Kurt**	**Prob***	**Lag order**	**ADF-stat**
Pre-C19	Intraday	P_t	37.560	42,7	13,7	-3,5	175,0	1,1	6,5	<0,01	33	-11,3
		WV_t	37.560	33,6	17,7	0	140,0	0,8	3,9	<0,01	33	-23,0
		T_t	37.560	111,6	66,9	-84	372,0	0,2	2,8	<0,01	33	-9,4
		$VarP24_t$	37.537	98,9	108,2	7,2	1847,7	6,1	63,0	<0,01	33	-16,1
Post-C19	Intraday	P_t	2.855	27.5	12,8	-79,2	90,0	-0,6	10,7	<0,01	14	-10,1
		WV_t	2.855	31.3	17,1	0,0	100,0	0,8	3,6	<0,01	14	-8,3
		T_t	2.855	151.4	64,1	-37,0	336,0	-0,4	3,1	<0,01	14	-5,1
		$VarP24_t$	2.832	75,6	112,9	7,2	1038,8	5,8	40,2	<0,01	14	-5,6
Pre-C19	Daily	P_t	1.565	42,7	9,7	17,1	89,0	0,9	4,4	0,038	11	-3,5
		WV_t	1.565	33,6	13,6	10,0	97,0	1,0	4,4	<0,01	1	-15,8
		T_t	1.565	111,6	60,5	-66,0	297,0	0,0	2,5	<0,01	1	-5,4
Post-C19	Daily	P_t	50	19,8	6,2	-5,5	26,0	-2,8	11,7	0,0704	3	-3,4
		WV_t	50	34,3	14,2	14	74,0	0,9	3,4	0,0224	3	-3,9
		T_t	50	110,5	41,2	38	192,0	-0,0	2,0	0,0888	3	-3,2

*From Augmented Dickey-Fuller – MacKinnon (1996) p-values at tested lag order

Where 'obs' stands for number of observations, 'Stdev' stands for standard deviation, 'Min' for minimal value measured and 'Max' for maximum value measured. 'Skew' stands for skewness, 'Kurt' for kurtosis.

From the descriptive statistics in Table 1 we can observe that during the C19-period the mean price is lower than on average between in the Pre-C19 period. As a proxy for volatility, we use volatility in the highest frequency (intraday), and we observe a slightly lower standard deviation during the Post-C19 period. This implies that from an *absolute* perspective we do not observe higher volatility, but when one considers prices are significantly lower in the Post-C19, we can say that *relatively* volatility is higher. Furthermore, is the lowest minimum price measured during the C19 period.

Skewness measures the amount of asymmetry of the data distribution, the closer to zero the more symmetric the distribution of the data is. Negative skewness implies a longer left-tail, concentration of data on the left-hand side of the histogram. Positive skewness implies the same but a longer right-tail. Bulmer (1979) suggests as a rule of thumb that skewness out-of the range -1 until +1 implies highly skewed data. This is the case for P_t during the Pre-C19 sample. Surprisingly, there is less skewness considering the prices variable P_t in the intraday Pre-C19 period than in the Post-C19 period, and the sign has changed implying that during the C19-period prices concentrated on the left-hand side of the distribution function instead of on the right-hand side during the Pre-C19 period. The sign change is the same in the daily time-frame, although the effect is much stronger, as we observe a skewness-statistic of 0,9 during the Pre-C19 period, and a skewness-statistic of -2,8 during the Post-C19 period.

Fatness of the tails is measured by the kurtosis-statistic. High kurtosis implies many extreme observations compared to a low kurtosis-measure. Kurtosis of exactly 3 is called mesokurtic and implies perfect normality. Distributions which result in values above 3 are *leptokurtic*, implying longer and fatter tails. Distributions that result in a kurtosis value below 3 are called *platykurtic*, implying short and thin tails compared to a normal distribution.

In the intraday data-frequency, kurtosis is high for prices when compared to windvelocities and temperature, implying many extreme values which we would expect given the sometimes spikey behavior of Day-Ahead prices. Windvelocities and temperature come close to values of 3 and therefore don't have extreme kurtosis values. Our price variable does show rather high kurtosis values, strongly suggesting a leptokurtic distribution. In the Post-C19 period the kurtosis is higher than during the Pre-C19 period implying more extreme values which is line with expectations and confirms our suspicion of the C19-period to be classified as a crisis-period.

Judging from these statistics we can conclude data is not perfectly normally distributed.

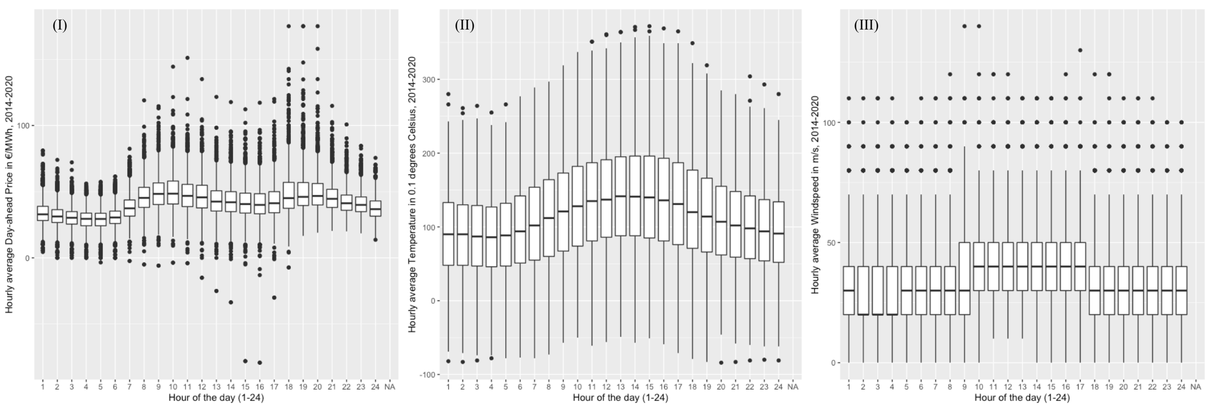

Figure 1: Boxplots of intraday prices (I), wind (II) and temperature (III).

In Figure 1 the boxplots of the intraday prices, wind velocities and temperatures are plotted. The strongest pattern is visible in the Day-Ahead prices. Electricity Day-Ahead prices are known to follow a certain pattern, as they are in general lower at night and higher during the day given demand fluctuations. Intraday electricity demand is known to follow a certain pattern, with a morning- and evening-peak and a night- and day lean (Guthrie and Videbeck 2007), which we find as well looking at the boxplot of intraday prices as depicted in Figure 1.I.

From studying the characteristics of the intraday data, modeling performance can be increased when one properly accounts for the characteristics of mean reversion, time of day effects, seasonal effects (summer and winter-effect plotted in Figure 2), time-varying volatility and extreme values. As the Day-Ahead prices do not trade during weekends, we have no weekend-data (and therefore weekend-price effects are not considered). Because prices can become negative, and because we do not model price changes, using log prices is deemed inappropriate.

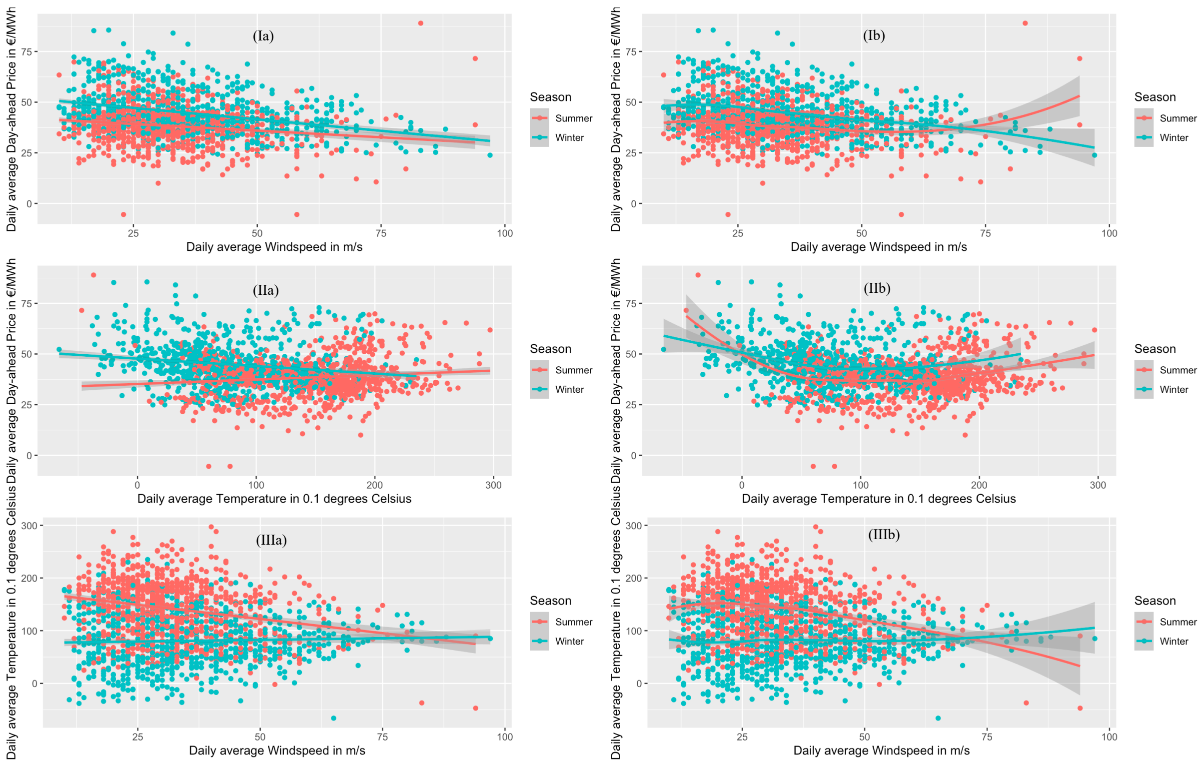

Figure 2: Scatterplot of the relations between daily average Day-Ahead prices and daily average windvelocities (I), daily average Day-Ahead prices and daily average temperatures (II) and daily average temperatures and windvelocities (III) of the full sample of daily prices (2014-2020). The data has been categorized between summer and winter observations. On the left hand-side a linear regression model has been fitted to the data. On the right-hand side a Kernel quadratic fit has been fitted in order to assess the data from a non-linear point of view.

3.3 Tests

Stationarity is a prerequisite for modeling financial time-series data. In order to assess if data is considered stationary an Augmented Dickey-Fuller Unit Root test is performed. The null-hypothesis of a unit-root is rejected in all series at the 10%-level which indicates that all data is stationary up until a certain lag, although we find higher p-values in the daily data-frequency, especially in the Post-C19 period (which can be expected given the lower data-frequency). If data is non-stationary, it can be accommodated in the ARIMA-analysis by differencing the data to become I(1) or even beyond.

For the GARCH-analysis the residuals of the best fitting ARIMA models are evaluated and discussed in the Results-section, as prerequisites for GARCH-models are more extensive than for ARIMA models, namely volatility clustering, leptokurtosis, autocorrelation and heteroskedasticity in the residuals. The tests performed are shortly explained in the section below.

The Jarque-Bera (JB) test is a test for normality. It determines if a sample has an excess kurtosis and a skewness equal to 0 or not. Thick tails distributions have a positive excess kurtosis (a higher peak), implying a leptokurtic distribution. The hypothesis is as follows for the JB Test:
H0: S = EK = 0
H1: S$\neq 0 \; or \; EK \neq 0$
The JB test rejects the null of the value exceeds the critical value provided by the Chi-squared (with 2 degrees of freedom). A high value of the JB test statistic is a sign is an indication of non-normality, and a low value indicates normality.

The Ljung-Box test tests model failures. It determines the degree of autocorrelation for a given lag length, often used to test for autocorrelation within the (standardized, squared) residuals.
H0: $\rho_1 = \rho_2 = \cdots = \rho_k$
H1:$\geq one \; of \; \rho_k \neq 0$
H denotes the number of lags up until a certain point. The Ljung-Box test on squared standardized residuals evaluates the dependence of the second moments with a time lag.

The ARCH-LM test tests if the data contains variation in the conditional volatilities, and thereby tests if (G)ARCH-effects are present in the data by regressing the squared error returns. When the null is rejected, the variance contains no homoskedasticity.
H0: $\propto_0 = \propto_1 = \cdots = \propto_n$
H1: $\geq one \; of \; \propto_n \neq 0$
If the ARCH-LM-test statistic is significant, it indicates there is an ARCH-effect present in the standardized residuals.

3.4 Distribution

The normal distribution, often referred to as the Gaussian distribution or bell-curve, is most commonly used in statistics given its simplicity, and that it can be mathematically integrated and differentiated indefinitely. This simplicity often provides an incentive to model using normal distribution, while the assumption of normality may in practice be violated. Electricity markets are known to fit another type of distribution better, namely the Student-t distribution (Huisman, 2009). The Student-t distribution adds a control parameter for fatness of tails, in the form of degrees of freedom.

In general, the normal-distribution function $f(x)$ can be described as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Where x are the values of the time-series, The Student-t distribution can be described as:

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma(\frac{v}{2})}(1 + \frac{x^2}{v})^{-(\frac{v+1}{2})}$$

Where v stands for the degrees of freedom and is the tail parameter, Γ stands for the Gamma function.

3.5 Models

3.5.1 ARIMA-Type models

The ARMA(p,q)-type takes into account the random nature and time correlations (Weron, 2014). The current value of price X_t is expressed linearly in terms of its p past values, the autoregressive part AR(p) and in terms of q previous values of the noise, the moving average part MA(q):

$$\phi(B)X_t = \mathrm{c} + \theta(B)\varepsilon_t, \qquad with\ \varepsilon_t \sim N[0, \sigma^2]$$

Where c is a constant, B is the backshift operator such that $B^h X_t \equiv X_{t-h}$ while $\phi(B)$ stands for $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$. Furthermore does $\theta(B)$ stand for $\theta(B) = 1 + \theta_1 B + \cdots + \theta_p B^q$, θ_p is the coefficient for the $AR(p)$ part, and B^q stands for the $MA(q)$ part. The error term ε_t finally has the condition of zero mean and finite variance. When $q = 0$ the model follows an $AR(p)$ process, and when p=0 the model follows an $MA(q)$ process.

When the dataset contains a unit-root problem and the stationarity assumption is therefore violated, the model should be differenced to follow an $I(I)$ process to overcome the unit-root problem. The ARMA model is then extended to become an $ARIMA(p, d, q)$ model as introduced by Box and Jenkins (1976). The model can then be rewritten as follows:

$$\phi(B)\nabla^d X_t = \mathrm{c} + \theta(B)\varepsilon_t, \qquad with\ \varepsilon_t \sim N[0, \sigma^2]$$

Where $\nabla_h x_t \equiv (1 - B^h)x_t \equiv x_t - x_{t-h}$ is the lagged$-h$ differencing operator. When $d = 0$ (and the data is stationary) the model follows an $ARMA(p, q)$ process. Or, written more extensively and intuitive for implementation of coefficients:

$$X_t = \mathrm{c} + \phi_t y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

The identification and estimation of all ARIMA-type models (so including the extensions) is performed by use of Maximum Likelihood.

3.5.2 ARIMA-Extensions: SARIMA

A Seasonal Autoregressive Integrated Moving Average, denoted as SARIMA, is an extension of $ARIMA(p, d, q)$. It adds three new hyperparameters to specify the AR, differencing (I) and Moving Average (MA). The trend elements are:

P	Seasonal autoregressive order
D	Seasonal difference order
Q	Seasonal moving average order
m	The number of time steps for a single seasonal period

All together the SARIMA model becomes specified as $SARIMA(p, d, q)(P, D, Q)m$. Intuitively we would expect adding seasonal components to the highly seasonal electricity prices makes for better forecasting accuracy.

Electricity prices are known to be heavily influenced by supply, which in itself is changing heavily with regards to the supply's source. These sources increasingly generate power from wind and solar sources, wind especially in the Netherlands. Adding these exogenous variables into our model intuitively therefore makes sense.

3.5.3 ARIMA-Extensions: ARIMAX

The method is analogous for all ARMA-type models. In this model, the current value of the spot price x_t is expressed linearly in terms of its past values, past noise and present and past values of our exogenous variables windspeed and temperature. Our ARMAX model can be generally specified as:

$$\phi(B)x_t = \theta(B)\varepsilon_t + \sum_{i=1}^{k} \psi^i(B)V_t^{(i)}$$

With $k = 1{,}2$ being the number of exogenous variables V and $\psi^i(B)$ is a compact notation form for $\psi^i(B) = \psi_0^i + \psi_1^i B + \cdots + \psi_{r_i}^i B^{r_i}$.

The ARMAX model in the transfer function' form is a bit more intuitive:

$$x_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t + \sum_{i=1}^{k} \tilde{\psi}^i(B) V_t^{(i)}$$

Where $\tilde{\psi}^i$ are the coefficient polynomials. $\theta(B) \equiv 1$ yields the dynamic regression form of the ARX-type model. Estimation of ARX-type models is often done with OLS-analysis or instrumental variables. Calibration of ARMAX coefficients is best done with maximum likelihood methods.

Knittel & Roberts (2005) use for intraday a rather simple model. They use:

$$p_t = \propto_t + \eta_t$$
$$\beta(L)\eta_t = \delta(L)\varepsilon_t$$

Where $\propto_t$ captures interaction terms per dummy for peak & off-peak, weekend and season and temperature squared and cubed to capture the nonlinearity (doesn't add to forecasting accuracy though). Knittel and Roberts continue their analysis using the same lags for AR and MA (1,24 and 25). They find the forecasting accuracy (measured by RMS forecasting error) to be the lowest of all models considered but find the temperature relationship to weaken during crisis-period.

Temperature has been used often in the literature. Given the increased share of renewables, in particular the increased share of wind resources in the energy mix in the last decade (Figure 13), and the expected growth of these resources in the coming years, we use these two datatypes as exogenous regressors in the ARIMAX-type model. We search for further patterns in the OLS-section.

3.5.4 ARIMA-GARCH models

When the residuals of the best fitting ARIMA model still show some volatility clustering, modeling the residuals as a GARCH-type model will likely increase forecasting accuracy. Because the linear ARIMA model does not capture volatility clustering as AR-type models assume homoskedasticity. With the ARIMA-model the mean effect will be captured. Moving forward the focus is solely on the volatility modeling, where the (G)ARCH-component can capture some non-linearity's of the non-constant conditional variance. In order to obtain the residuals of our ARIMA-model, we compute:

$$r_t = p_t - y_t$$

$$x_t = \mu + \sum_{t-i}^{m} \phi_i x_{t-i} + \sum_{t-j}^{n} \theta_j \varepsilon_{t-j} + \varepsilon_t$$
$$\varepsilon_t = \sigma_t z_t \qquad z_t \sim N(0,1)$$

Engle's (1982) AutoRegressive Conditional Heteroskedastic (ARCH) model was the first model that successfully addressed heteroskedasticity, by modeling the conditional variance of a time-series as an AR-process, the weighted sum of squared preceding observations. The ARCH(q) conditional variance can be written in general as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \propto_i \varepsilon_{t-j}^2$$

The standard GARCH(q,p) model, following Bollersev (1986) may be written as:

$$\sigma_t^2 = \left(\omega + \sum_{j=1}^{m} \zeta_j\, v_{jt}\right) + \sum_{j=1}^{q} \propto_i \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j\, \sigma_{t-j}^2$$

σ_t^2 denotes the conditional variance, ω the intercept, ε_{t-j}^2 the residuals from the mean filtration process of the ARIMA-model in our case. The optional m number of external regressors v_{jt} are pre-lagged.

For the GARCH(1,1)-type our variance model of our standard univariate GARCH model, the variance model becomes:

$$\sigma_t^2 = \omega + \propto \varepsilon_{t-j}^2 + \beta_j \sigma_{t-j}^2$$

The mean model, uses the residuals of a certain ARIMA-model. When residuals are prove to be leptokurtic, the distribution of the model can be best specified as a Student-t distribution ('std' in the statistical software package). The Student-t model looks slightly different, but the interpretation will be similar to the normal distribution:

$$x_t = \mu + \theta_1 (r_{t-1} - \mu) + \theta_2 \varepsilon_{t-1} + \varepsilon_t$$
$$\varepsilon_t = \sigma_t z_t \qquad z_t \sim N(0,1)$$

The steps performed in the analysis are after specifying the mean equation as an ARMA-model and specifying and estimating the GARCH-model, we will use the model to make forecasts. We will test limit our models to a GARCH(1,1) and an ARCH(2) model in order to compare the performances. As described in the literature review, we expect the GARCH method to especially work well during the crisis period and performs by nature not well for short-term forecasting except when combined with ARIMA where the residuals of the regression part are modelled further with a GARCH-process.

The identification and estimation of GARCH models is performed by use of Maximum Likelihood.

3.5.5 Regime-Switching models

Markov-Switching models allow parameters to vary with time. Switching between states or regimes can be either abrupt or smooth. Dynamic regression models. are best applied in high-frequency data as they allow for quick adjustments after the process changes state. Autoregressive models are used for our daily data as they allow for gradual adjustments after the process changes state.

As Huisman (2007) points out is a drawback of regime switching models that the probability that a spike occurs is constant over time. Mount er al (2006) proposes a more advanced regime switching model in which spike occurrence has a time varying probability and depends on reserve margin. these problems by using dummy's for weekend and weekdays and using four seasons. Weather conditions are still assumed to be constant over time. This is not realistic since spikes can occur as a result of demand or supply shocks. Mount, Ning and Cai (2006) argue that a spike depends on the reserve margin. Huisman (2008) proposes daily temperature as a variable that influences the probability on a spike. They model two regimes: Regime 1 reflects a normally functioning market and 2 reflects a non-normal market due to a shock in demand and/or supply that results in a spike. Huisman argues temperature affects demand, as in cold winter days demand spikes and during summer air-conditions are turned on which also can cause spikes. Today, when renewables represent an increasing part of the energy mix, supply is affected as well. Huisman finds in summer that a higher-than-expected temperature leads to a higher probability of a spike, and during winter a lower-than-expected temperature leads to a higher probability of a spike. Therefore, including temperature and wind speeds poses for interesting research.

To model the regime switching model, we follow Hamilton (1989, 2005), Mount et al. (2006) and Huisman (2008):

Regime 1 is the normal market condition regime, which follows a mean-reverting process:

$$\partial x_t = x_{t-1} - \propto x_{t-1} + \sigma_1 \varepsilon_{1,t}\,, \qquad with\ \varepsilon_{1,t} \sim N[0,1]$$

Regime 2 is the spike regime:

$$\partial x_t = \mu_2 + \sigma_2 \varepsilon_{2,t}, \qquad with\ \varepsilon_{2,t} \sim N[0,1]$$

Let $p_{t,i,j}$ be the transition probability of moving from regime j on day t-1 to regime i on day t. The transition probabilities are assumed to be a function of temperature. The transition probability $P_{t,2,1}$ can be described as:

$$P_{t,2,1} = \lambda_1 + \lambda_2 S_t E_{t-1}[\lambda_3(1 - S_t)] E_{t-1}(\Delta\tau_{t,h})$$

The other transition probabilities are:

$P_{t,1,1} = 1 - P_{t,2,1}$

$P_{t,2,2} = \lambda_4$

$P_{t,1,2} = 1 - P_{t,2,2}$

Following Huisman and Mahieu (2003), Huisman (2007) and Mount, Ning and Cai (2006) we will in order to preserve values between 0 and 1, apply a logistical transformation for the actual probability P^*:

$$P^* = \frac{e^{p_{t,2,1}}}{1 + e^{p_{t,2,1}}}$$

Or following Huisman (2008), who uses a framework where the natural log of the Day-Ahead price s_t depends on a stochastic component x_t and a deterministic component d_t which accounts for a predictable component of Day-Ahead electricity prices such as the mean-price level and seasonality (by means of usage of dummy variables).

$$s_t = d_t + x_t$$
$$d_t = \mu + \beta_1 w_t$$

Furthermore, can we elaborate on the AR-GARCH models. A limitation of GARCH-type modeling is that estimates of GARCH-models can be biased by structural breaks in the volatility dynamics. Markov-switching GARCH models (MSGARCH) could improve results, as it as explained above allows for two states. We obtain similar GARCH-type models for the conditional variance, but with k regimes depending on the state s_t (Haas et al., 2004):

$$\sigma_{1,t}^2 \equiv \omega_1 + \propto_1 \varepsilon_{t-j}^2 + \beta_1 \sigma_{1,t-j}^2$$
$$\sigma_{K,t}^2 \equiv \omega_K + \propto_K \varepsilon_{t-j}^2 + \beta_K \sigma_{K,t-j}^2$$

The discrete state variable s_t changes according to a first-order Markov chain with transition matrix P. We use Maximum Likelihood to estimate the model. We will use a two-state Markov-Switching GARCH model with a GARCH(1,1) type model following a Student-t distribution.

3.5.6 OLS-Extension

This section of the paper can be viewed merely as an extension and lays more focus on the economic implication of the changing power market instead of forecasting purposes. The share of renewable energy in the final energy consumption is plotted in Figure 12. We observe an increase in wind energy consumption from 1,04% in 2014 to 1,70% in 2018, while the solar produced consumption increased from 0,19% in 2014 to 0,60% in 2018. As the share of renewables and more specifically electricity generated by wind sources has grown strongly, one might expect extreme windspeeds to have a more significant influence on prices and volatility, given that wind sources are more volatile than sources such as gas and especially coal (Figure 5), and oversupply (as electricity is in general) is hardly storable.

Model 1:

For the first model, we set up a simple OLS-model where we test the influence of wind by dividing windspeeds into low-to-high categories ('quantiles') and test them by use of dummy variables.

$$P_t = \alpha + \beta_1(d[0-19]WV_t) + \beta_2(d[20-39]WV_t) + \beta_3(d[40-59]WV_t) + \beta_4(d[60-79]WV_t) + \beta_5(d[80-99]WV_t) + \beta_6(d[100-119]WV_t) + \beta_7(d[>120]WV_t) + \varepsilon_t$$

were $d[0-19]WV$ indicates the dummy variable that is equal to one during periods in which windvelocities were measured between 0 and 19 m/s.

Model 2:

In addition to the previous model, we want to assess basically the same two hypotheses, in order to find out the relation windspeeds have with (a 24-hourly rolling) volatility and how this relationship evolves over time as the share of wind increases.

$$VAR24_t = \alpha + \beta_1(d[0-19]WV_t) + \beta_2(d[20-39]WV_t) + \beta_3(d[40-59]WV_t) + \beta_4(d[60-79]WV_t) + \beta_5(d[80-99]WV_t) + \beta_6(d[100-119]WV_t) + \beta_7(d[>120]WV_t) + \varepsilon_t$$

Model 3:

Similar to the analysis we performed in Models 1 & 2, it proves interesting to assess the relationship between temperature and prices. We again created quantiles for the dummy variables from low to high. We furthermore created a variable that assesses high positive and negative deviations from the average for that month and the average of hours, in order to assess prices when temperature is in fact more extreme than can be considered 'normal'. Characteristics of temperatures are that they tend to deviate strongly between intraday hours and among months. Therefore, variables dTL_t and dTH_t were created that took both of those seasonal problems into account, and purely focused on relative outliers, if measured temperature was higher or lower than the average mean value of temperature T at time per month, per hour plus or minus the (24-hourly rolling) average standard deviation at time per month m, per hour h. Given normal distribution we thereby capture circa. 34% of the most extreme temperatures:

$$dTH_t = 1\ if: T_t < \sum_{m}^{12}\sum_{h}^{24}(\bar{T}_{m,h} - (1 * \sigma 24_{m,h}))$$

$$dTH_t = 1\ if: T_t > \sum_{m}^{12}\sum_{h}^{24}(\bar{T}_{m,h} + (1 * \sigma 24_{m,h}))$$

Where $\bar{T}_{m,h}$ indicates the average temperature per month m per hour h, and $\sigma 24$ denotes the trailing standard deviation of temperatures of the current observation and the previous 23 observations.

The model, with lagged prices added as a control variable, is set up as follows:

$$P_t = \alpha + \beta_1 P_{t-1} + \beta_2 T_t + \beta_3 dTH_t + \beta_4 dTL + +\beta_5(dTL * Temp_t) + \beta_6(dTH_t * Temp_t) + \varepsilon_t$$

Here the dummy variable will show prices during extreme temperatures, while the interaction term will also take into account the change in the slope. Analytically this is explained below for dTH_t:

$$\frac{\partial E(P_t | dTH_t = 0)}{\partial Temp_t} = \beta_2$$

$$\frac{\partial E(P_t | dTH_t = 1)}{\partial Temp_t} = \beta_2 + \beta_6$$

3.6 Model Performance

3.6.1 Information criteria

When judging which ARIMA(p,d,q) process fits the data best, we will determine analytically from the ACF and PACF plots which lagged variables will fit the data best for the AR(p) process. By analyzing the ACF and PACF plots, the correlation among the residuals can be examined to check for autocorrelation. Furthermore, will the data be tested on stationarity. For the I(d) process we will judge from stationarity tests if the data has a unit-root problem and therefore needs to be differenced to follow an I(1) or possibly I(2) process. In order to determine which (S)AR(I)MA(X) will fit the data best, we use the Akaike Information Criterion (AIC). These criteria are calculated as follows:

$$l = -n \ln(\sigma) + \sum_{i} \ln\,[\phi(z_i)]$$

$$AIC = BIC = -2l + 2kn_{par}$$

Where l stands for the likelihood function, n_{par} the fitted number of parameters and k is the penalty per parameter used which differs among AIC and BIC, the default is $k = 2$ for AIC and k=log(n) for BIC. AIC penalizes the inclusion of modeling additional variables, where BIC is a variant with a stronger penalty for including additional variables. Minimal AIC and BIC values are used as model selection criteria. A limitation of purely assessing the AIC and BIC criteria is that there may be a better fit by an ARIMA model with more AR or MA lags added, but that fit is not worth the loss in parsimony imposed by the addition of these lags. When judging which model fits best, we focus on AIC for the sake of simplicity. Log-Likelihood is also reported but this metric is only evaluated when in doubt given the other metrics.

3.6.2 Evaluating models performance

Most widely used in electricity price forecasting are point forecasts and so its evaluation. Most widely used measures of accuracy are based on absolute errors (AE).

$$MAE = \frac{\sum \left| P_h - \hat{P}_h \right|^2}{n}$$

P_h the actual price, $\hat{P}_h$ the forecasted price in period h, n the number of fitted values. Knittel and Roberts (2005) note that given the stochastic nature of electricity prices the forecast error, the root mean square errors (RMSE) will be necessarily high. It is calculated as the square root of the average of squared differences between predicted and actual values. The change in the error will therefore be a better indicator for judging out-of-sample forecasting performance. Additionally, important to note for the RMSE is influenced by changes in mean and variance across models.

$$RMSE = \sqrt{\sum_{h=1} (P_h - \hat{P}_h)^2}$$

When judging the models' performance, a combination of the above explained point forecast evaluation metrics will be used, but most focus will lay on RMSE for most models, MAE will be reported to judge from when in doubt. The difference between the two metrics is in that for RMSE, the errors are squared prior to being averaged, giving high weights to large errors. Given that large errors are an important indicator we will use this metric as most important performance measure. Given that the MAE uses an absolute value, the interpretation of the metric is a little more advanced since positive and negative errors will cancel out. Interpretation wise, the lower both values are the better the performance of the predicted values versus observed.

3.7 Expected Outcomes

In this section the hypotheses on which this thesis will be built will be explained and the motives behind them. Moreover, will the expected outcomes be discussed in this section.

We would expect to find improvement when modeling when a more complex model that incorporates more of the power market's electricity. By not combining the models, we will isolate which model performs best per period. Thereby we are able to reason which characteristic performs best per period and the (economic) motive behind it which will add to our understanding of the dynamic of the market. We expect a SARIMA and ARIMAX to outperform the same ARIMA-model, and the ARIMA-GARCH model to outperform those. Especially during the crisis-period we expect ARIMA-type models to perform less well given stronger volatility-clustering. The ARIMA-GARCH models should therefore show a stronger increase in performance compared to ARIMA-type models given the higher volatility than during the crisis-period. Furthermore, we expect better performance when the different distributions are used, if there are indications that residuals are not normally distributed.
We would expect the intraday models to achieve a higher modeling accuracy, given the more observations and therefore higher ability to learn from past data. On the other hand, will autocorrelation likely be higher and therefore will the models depend on more lags, complicating computations. Daily data however are interesting as the lower frequency of the data also provides perspective to forecast for a longer-term horizon into the future, as $t + 1$ is one Day-Ahead in daily frequency, and one hour ahead in the intraday frequency. Also will white-noise in the residuals be relatively easier to achieve for these type of models given the lower frequency. Interesting to analyze will be the difference in performance among not only the performance of ARIMA-GARCH during the crisis-period, but as well the difference in-between the data-frequencies. The higher the frequency, the better the proxy of volatility is that will be assessed. Comparing the with respect to GARCH models periods will be done by using the weighted ARCH-LM test on the respective models.

For the purpose of this study, we want to examine the dynamics of the Dutch Day-Ahead market, and the change of these dynamics during the C19 crisis-period and in between daily average price dynamics and intraday dynamics. This will be realized by fitting the data to a wide-range of models in order to assess which ones model this market best in the normal period and in the C19 crisis-period. Furthermore, will we assess the influence of exogenous regressors windvelocities and temperature, and if this relation has become stronger over the years.

This study follows the research question: "Which model-types show the best in-sample performance, and model the dynamics of the Dutch Day-Ahead market best, and how does the Pre-C19 period compare to the Post-C19 period in terms of these criteria?"

In the OLS-analysis, we are curious to find out if indeed the expectation of higher influence of wind correlates with lower prices, and if this pattern becomes stronger over-time as the share of wind in the energy-mix grows. We do work with hypotheses for this section, given the different nature of this type of analysis:

$H_{0A1}: There\ is\ no\ relation\ between\ windspeeds\ and\ prices$
$H_{1A1}: There\ is\ a\ relation\ between\ windspeeds\ prices$

$H_{0A2}: The\ relation\ is\ not\ necessariliy\ becoming\ stronger\ in\ recent\ years$
$H_{1A2}: The\ relation\ is\ becoming\ stronger\ in\ recent\ years$

$H_{0B1}: There\ is\ no\ relation\ between\ windspeeds\ and\ extreme\ (rolling)\ price\ volatility$
$H_{1B1} There\ is\ a\ relation\ between\ windspeeds\ and\ extreme\ (rolling)\ price\ volatility$

$H_{0B2}: The\ relation\ is\ not\ necessariliy\ becoming\ stronger\ in\ recent\ years$
$H_{1B2}: The\ relation\ is\ becoming\ stronger\ in\ recent\ years$

$H_{0C1}: There\ is\ no\ relation\ between\ prices\ and\ extreme\ temperatures$
$H_{1C1}: There\ is\ a\ relation\ between\ prices\ and\ extreme\ temperatures$

$H_{0C2}: The\ relation\ is\ not\ necessariliy\ becoming\ stronger\ in\ recent\ years$
$H_{1C2}: The\ relation\ is\ becoming\ stronger\ in\ recent\ years$

4. Results

4.1 Intraday – PreC19

4.1.1 ARIMA-type models

The ARIMA-type models for the prices on intraday frequency[3] prices, show high autocorrelation and therefore in-sample performance keeps on improving when more AR-lags are added. From the ACF & PACF plots as displayed in Figure 7, the autocorrelations are statistically significant far beyond 100 lags. It appears that around the 24th/25th lag from the PACF a change seems to take place which makes sense due to the 24 hour-structure of prices. The pattern appears to repeat itself as it diminishes. Therefore, AR-lags of 24 are used, as well as an easier model with only an AR-lag of 1 for comparison. We found based on in-sample information criteria that an AR-lag of 3 was optimal. Given that the data is stationary (Table 1), differencing the series will not be optimal for lower $AR(p)$ & $MA(q)$ lags and therefore I(0) should be optimal.

[3] Due to computational difficulties the hourly Pre-C19 dataset has been shrunk so it holds data from 2016-2019

Our expectation was that simple ARIMA models would perform less well during the Post-C19 period given the stronger volatility-clustering which is more difficult for a simple ARIMA model to capture. Based on the in-sample performance metrics, this is difficult to state. We do not necessarily find evidence in our models of this expectation.

Table 2: In-sample goodness-of-fit tests AIC & Log-Likelihood, intraday frequency

	Pre-C19		**Post-C19**	
Model	**AIC**	**LogLik**	**AIC**	**LogLik**
ARIMA(1,0,3)	161.210	-80.602	17.099	-8.547
ARIMA(24,0,3)	151.197	-75.570	18.301	-7.677
SARIMA(1,0,3)(3,0,1)[24]	147.355	-73.669	15.230	-7.605
SARIMA(24,0,3)(3,1,1)[24]	**147.148**	**-73.543**	**15.046**	**-7.492**
ARIMAX(1,0,3)	159.634	-79.809	16.371	-8.178
ARIMAX(24,0,3)	150.965	-75.452	15.296	-7.616
ARMA(24,3)-GARCH(1,1)norm	147.268	-70.888	14.441	-7.201
ARMA(24,3)-GARCH(1,1)std	142.480	-68.580	14.081	-7.061
ARMA(24,3)-GARCH(1,1)sstd	142.447	-68.563	14.082	-7.061
ARMA(24,3)-ARCH(2)norm	143.972	-75.601	14.568	-7.716
ARMA(24,3)-ARCH(2)std	133.727	-70.218	13.502	-7.211
ARMA(24,3)-ARCH(2)sstd	143.962	-70.212	14.443	-7.211
RSGARCH	257.641	-128.810	26.795	-13.388

The best performing models have been highlighted in **Bold** which will be evaluated in Section 6.3.

4.1.2 ARIMA-Extensions

Given the strong seasonal patterns as depicted the decomposed data in Figure 9, adding seasonal components has potential to add to the (in-sample) performance of simpler ARIMA models. We set $m = 24$ given the strong intraday patterns and the possibility of a forecasting horizon of $t + 24$ hours. We expect the exogenous variables can also add to the performance of the simpler ARMA models.

The SARIMA-type models are optimized for $(P, D, Q)m$ at (3,0,1) we found by a trial-and-error method. The SARIMA-type models indeed improve the simpler ARIMA-type models judging from the in-sample goodness-of-fit metrics. The relative improvement is found to be stronger for the information criteria of the SARIMA-type models during the Post-C19 period than in the Pre-C19 period. The performance metrics as displayed in Table 3 show strong improvements for the SARIMA-type models compared to their simpler variants in both the Pre-C19 and Post-C19 periods. Judging from the RMSE and MAE, the SARIMA-type models are even the best performing models of all tested models in the intraday frequency data during the Pre-C19 period and to a lesser degree in the Post-C19 period.

Adding exogenous regressors solar and wind, the ARIMAX-models' improvements are not or barely present surprisingly. The information criteria show only a marginal improvement in both periods. The performance measures in Table 3 also only report marginal better in-sample performance than their

simpler ARIMA-equivalents. In the intraday frequency, SARIMA-type models outperform ARIMAX-type models. There is no strong difference between the two periods.
In the Post-C19 period we again observe ARIMAX models underperforming, while SARIMA models outperform the simpler modes. We find no clear difference during the Pre- and Post-C19 period concerning the ARIMA-extensions.

4.1.3 ARIMA-GARCH models

As indicated in the first section of the results, we picked the ARMA(24,3)-type model for further modeling, for computational simplicity and to be able to capture the pure effect modeling the residuals as a GARCH-type model has. To match the assumptions required for GARCH-modeling, we first tested the residuals using the tests as explained in 5.3. The results of the residuals of the ARMA(24,3) and the best fitting ARMA-type model for the Daily-analysis are shown in Table 12. The residuals should exhibit heteroskedastic behavior, they should show signs of volatility clustering. In Figure 11 (Ia & IIa) the residuals of the ARMA(24,3) model are plotted, and show signs of volatility clustering. Also shown in this figure are the density plots (Ic & IIc), which indicates a fat-tails leptokurtic distribution.
The ARCH-LM test formally tests for ARCH-effects. The results indicate there is an ARCH-effect present in the residuals of the ARMA(24,3) model in both periods, which makes proceeding with modeling the residuals as a (G)ARCH-type process reasonable. The Ljung-Box test for autocorrelation shows mixed results. The null-hypothesis for autocorrelation is rejected in the Pre-C19 period and Post-C19 period, whereas we fail to reject the null-hypothesis for the Daily periods. Using robust standard errors can provide a solution when estimating the ARMA-GARCH model. The Jarque-Bera test indicates what we expected from the density plot of the residuals (Figure 11), the residuals are not normally distributed, but likely distributed in a leptokurtic way. This indicates that a normally distributed (G)ARCH-type model may not be optimal, a Student-t distribution may be a better fit.
From the weighted ARCH-LM test as displayed in Table 13, we reject the null-hypothesis in most periods given the low p-values, for each GARCH-type model and for each lag of 3, 5 and 7 and have to conclude there are still ARCH-effects remaining in the data and these residuals can not be considered White noise. An exception can be made for the case of the ARMA(24,3)-GARCH(1,1) model in the Pre-C19 period with a Student-t distribution, where we reject the null-hypothesis of remaining ARCH-effects.

Table 3: Performance measures per model for the intraday data, divided into pre and post C19 period

	Pre-C19		**Post-C19**	
Model	**RMSE**	**MAE**	**RMSE**	**MAE**
ARIMA(1,0,3)	6,06	4,02	4,83	3,29
ARIMA(24,0,3)	4,95	3,13	3,55	2,31
SARIMA(1,0,3)(3,0,1)[24]	4,60	2,87	3,43	2,18
SARIMA(24,0,3)(3,1,1)[24]	**4,57**	**2,85**	**3,37**	**2,16**
ARIMAX(1,0,3)	5,87	3,78	4,58	2,97
ARIMAX(24,0,3)	4,93	3,12	3,55	2,34
ARMA(24,3)-GARCH(1,1)norm	4,99	3,13	3,65	2,39
ARMA(24,3)-GARCH(1,1)std	5,01	3,10	3,64	2,33
ARMA(24,3)-GARCH(1,1)sstd	5,01	3,11	3,64	3,64
ARMA(24,3)-ARCH(2)norm	4,96	3,13	3,61	2,36
ARMA(24,3)-ARCH(2)std	5,02	3,09	3,68	2,33
ARMA(24,3)-ARCH(2)sstd	5,02	3,10	2,33	3,68

The best performing models have been highlighted in **Bold** which will be evaluated in Section 6.3.

Judging from the information criteria, we find the Student-t distribution to achieve a better AIC and Log-Likelihood, which again was to be expected given the fat-tails distribution.

Judging from the performance measures as displayed in Table 3, we find improved metrics RMSE and MAE for the GARCH-type models, although not as strongly improved as the SARIMA-type models in the Pre-C19 period. We find in the in-sample performance that indeed the ARMA-(G)ARCH models provide significant improvements. The information criteria show a more significant improvement over the simpler models than the performance measures RMSE and MAE surprisingly in both periods. We do not find a stronger performance of the GARCH-type models in the intraday data frequency during the Post-C19 period.

For both periods, in nearly all performance metrics, the SARIMA(24,3)(3,1,1)[24] is the best performing model. Only in the Post-C19 period the ARMAGARCH(2)'sstd' performs better in-sample, but the SARIMA model has a better MAE. Regarding information criteria, the ARMA-GARCH-type models outperform the SARIMA-type models. The SARIMA(24,3)(3,1,1)[24] model will be evaluated in Section 6.3.

4.1.4 GARCH Regime-Switching models

We find the regime switching GARCH models to underperform so strongly that we disregard them for the analysis. See Table 14 for the results.

4.1.5 OLS-Extension

Model 1

The results of Model 1 are summarized in Table 4 and show that in the lowest two windvelocitiy quantiles (0-19 and 20-39 m/s), prices are highest and significant in years 2015, 2016, 2017, 2019 and during the Post-C19 period. The only two periods in which this is not the case is in period 2014 and period 2018. The lowest coefficients are found in the categories of windspeeds of 80 and above, implying lower prices when windspeeds are high. The year 2014 behaves irregular, as the first two quantiles show the lowest coefficient, however they are insignificant. Furthermore, is the highest coefficient of 2014 found in the 100-119 windspeed quantile. If we move one category up in terms of windspeeds, we almost one-for-one observe a decreased coefficient, implying that there is a (negative) relation between windspeeds and heights of prices and the magnitude gets stronger in more extreme quantiles, leading to a rejection of H_{0A1}.

The model is set-up as follows:

$$P_t = \alpha + \beta_1(d[0-19]WV_t) + \beta_2(d[20-39]WV_t) + \beta_3(d[40-59]WV_t) + \beta_4(d[60-79]WV_t) + \beta_5(d[80-99]WV_t) + \beta_6(d[100-119]WV_t) + \beta_7(d[>120]WV_t) + \varepsilon_t$$

Table 4: Results OLS Model 1. OLS dummy regression for separate periods of the full intraday sample

	Period													
	2014		**2015**		**2016**		**2017**		**2018**		**2019**		**Post C-19**	
	Robust standard error regression, P_t as dependent variable													
Parameter	**Coeff.**	**Sign.**												
α	41.59	***	53	***	37.25	***	36.89	***	69.75	**	42.94	***	7.02	
β_1	-0.80		-12.24	**	-3.27	***	4.69	***	-13.48	**	0.28		23.82	***
β_2	0.97		-10.74	*	-1.65	***	5.25	***	-13.07	***	1.65	**	22.54	***
β_3	2.04	*	-11.77	**	-4.71	***	3.11	***	16.51	***	-1.8	**	18.46	***
β_4	2.39	**	-13.25	**	-6.71	***	0.46		-20.92	***	-1.77	**	10.86	*
β_5	N.A.		-17.3	***	-10.84	***	-1.67		-22.24	***	N.A.	**	7.26	
β_6	6.88	*	-17.15	***	-11.59	***	0.345		N.A.		-0.79		N.A.	
β_7	N.A.		N.A.	***	N.A.		N.A.		-22.88	***	N.A.		N.A.	
R-squared	0.0091		0.0167		0.0311		0.0148		0.0272		0.0193		0.0946	

*, **, *** correspond to significance at 10%, 5% and 1%

Where 'Coeff' stands for coefficient and 'Sign' for significance

Whether the influence of wind become stronger, as might be expected from the increased renewable production from 2014 onward, is implicitly tested. Making it therefore difficult to conclude, but the pattern of low wind correlating with high prices, and high wind correlating with lower prices does appear to become stronger as time moves on, and the pattern is strongest in the Post-C19 period. Although we do not find conclusive evidence, and correlation should not be confused with causation. Also is the Post-

C19 period a much shorter period than the other considered periods. Given the limited evidence, we fail to reject H_{0A2}.

Model 2

Similar to Model 1, we set up a model that instead of the pure prices, tests the 24-hourly rolling volatility as a dependent variable onto the same temperature quantiles as in Model 1. The model is set up as follows:

$$\begin{aligned} VAR24_t = \alpha &+ \beta_1(d[0-19]WV_t) + \beta_2(d[20-39]WV_t) + \beta_3(d[40-59]WV_t) \\ &+ \beta_4(d[60-79]WV_t) + \beta_5(d[80-99]WV_t) + \beta_6(d[100-119]WV_t) \\ &+ \beta_7(d[>120]WV_t) + \varepsilon_t \end{aligned}$$

Table 5: Results OLS Model 2. OLS dummy regression for separate periods of the full intraday sample

	Period													
	2014		**2015**		**2016**		**2017**		**2018**		**2019**		**Post C-19**	
	Robust standard error regression, $VAR24_t$ as dependent variable													
Parameter	**Coeff**	**Sign**												
α	148.83	***	93.11	***	76.58		40.43	***	397.93	***	48.48	***	125.30	***
β_1	-79.37	***	5.15	***	14.58	***	96.93	***	-274.33	***	26.92	***	-45.51	*
β_2	-70.43	***	-4.77	***	11.17	***	84.98	***	-256.43	***	32.74	***	-52.29	**
β_3	-51.71	***	-4.59	***	-1.41		55.21	***	-260.14	***	23.64	***	-62.26	***
β_4	-21.43	***	7.50	***	-17.14	***	57.73	***	-259.20	***	22.42	***	-10.19	
β_5	N.A.		3.11	***	-23.29	***	57.25	***	-199.13	***	N.A.	***	20.18	
β_6	-29.59		-44.11	***	-18.95	***	18.51		N.A.	***	-17.84	***	N.A.	
β_7	N.A.		N.A.		N.A.		N.A.		-286.15	***	N.A.	***	N.A.	
R-squared	0.0990		0.0099		0.0126		0.0087		0.0198		0.0116		0.0183	

*, **, *** correspond to significance at 10%, 5% and 1%

Where 'Coeff' stands for coefficient and 'Sign' for significance

The intercept itself says a lot about volatility in the given period, in contrast to the intercept in Model 1. It shows that volatility was not even highest during the Post-C19 period, but in 2018 and 2014 the Day-Ahead market was more volatile. The pattern of volatility in relation to wind in general is less clear than it was for prices in Model 1. In 2014 and in the Post-C19 period we observe an increasing pattern between higher windspeeds and higher volatility. In 2016 and 2017, we observe the highest volatility in the lowest windspeed quartile, and the lowest volatility in the highest quantile, where perhaps one would expect more volatility during extreme weather. In more recent periods 2019 and in the Post-C19 period, we do observe lowest volatility in the lowest quantile and highest volatility in the highest quantile (although insignificant in the Post-C19 period).

We find the relation not strong enough to conclude that we can reject the H_{0B1} nor do we find enough evidence that we clearly observe a stronger influence of wind onto the variance of prices as time moves on and we fail to reject H_{0B2}.

A limitation of this way of testing is that the volatility is tracking 23 previous and the current price observations, and this is yesterday's volatility and the observed price is set today and agreed for delivery tomorrow, a limitation of the Day-Ahead market itself. However, may this data still be indicative of a relationship although perfectly testing is rather difficult.

Model 3

The results of our third model, are summarized in Table 6. In the Full-Period we find our general results, thereafter we will examine if these general observations are in line when the separate years are analyzed. The model looked as follows:

$$P_t = \alpha + \beta_1 P_{t-1} + \beta_2 T_t + \beta_3 dTH_t + \beta_4 dTL + +\beta_5(dTH_t * T_t) + \beta_6(dTL * T_t) + \varepsilon_t$$

Table 6: Results OLS Model 3. OLS dummy regression for separate periods of the full intraday sample

	Full-Period		**2014**		**2015**		**2016**		**2017**		**2018**		**2019**		**C19**	
	Robust standard error regression, P_t as dependent variable															
Parameter	**Coeff**	**Sign.**														
α	3,720	***	7,410	***	6,738	***	3,275	***	5,560	***	6,153	***	4,818	***	0,626	
St dev	*-0,170*		*0,450*		*0,381*		*0,403*		*0,550*		*0,618*		*0,387*		*0,466*	
β_1	0,907	***	0,840	***	0,829	***	0,892	***	0,871	***	0,860	***	0,897	***	0,912	***
St dev	*0,004*		*0,009*		*0,009*		*0,013*		*0,012*		*0,014*		*0,008*		*0,020*	
β_2	0,001		-0,004	*	0,003	*	0,002		-0,003	**	0,011	***	-0,004	***	0,012	***
St dev	*0,001*		*0,002*		*0,002*		*0,001*		*0,002*		*0,002*		*0,001*		*0,002*	
											-					
β_3	-0,733	***	-1,202	**	-1,405	***	-0,658		-0,326		0,804	*	-1,060	***	-0,487	
St dev	*0,164*		*0,473*		*0,462*		*0,414*		*0,473*		*0,473*		*0,364*		*1,112*	
β_4	0,562	***	-0,014		0,430		0,728	***	1,078	**	1,603	***	0,534	*	0,630	
St dev	*0,145*		*0,421*		*0,331*		*0,268*		*0,423*		*0,443*		*0,294*		*0,456*	
β_5	0,004	***	0,006	**	0,007	**	0,004		0,000		0,002		0,007	***	-0,001	
St dev	*0,001*		*0,003*		*0,003*		*0,003*		*0,003*		*0,003*		*0,002*		*0,005*	
											-					
β_6	-0,004	***	-0,001		0,000	*	-0,002		-0,010	**	0,012	***	-0,005	*	-0,001	
St dev	*0,001*		*0,004*		*0,003*		*0,003*		*0,004*		*0,004*		*0,003*		*0,003*	
R-squared	0,8278		0,7109		0,6987		0,8057		0,7866		0,765		0,8257		0,8587	

*, **, *** correspond to significance at 10%, 5% and 1%

Where 'Coeff' stands for coefficient and 'Sign' for significance

In the Full-Period we find the first lag of prices to be highly influential on current prices, which was to be expected given the high autocorrelation we found in our intraday electricity price dataset.

Temperature as a variable in itself appears not to have a strong effect onto current prices and is also the only variable in the Full-Period that is insignificant, no clear conclusions of the temperature in general onto prices can be drawn. When the temperature variable is split-up as explained in methodology into a relative measure of high and low temperatures, we find that the coefficient for the dummy variable which is effective during high relative temperatures (β_3), we find a strongly negative coefficient, implying the intercept for prices is lower during high relative temperatures. We find the intercept during relative low temperatures, coefficient for β_4, to be strongly positive, although less positive than β_3 was negative. When assessing the coefficients for interaction terms, β_5 & β_6, the slope of the regression is analyzed, as explained in the methodology. We observe β_5 to be (slightly) steeper upward and β_6 to be slightly steeper downward although the effects are small. Thereby we can conclude that in the Full-Period, during high relative temperatures the intercept of Prices is strongly more negative, and the slope of the regression is slightly more positive compared to the overall slope to be slightly steeper of the regression regarding temperatures. During low relative temperatures, the effect is opposite but a bit less strong. Therefore, we reject H_{0C1} and accept H_{1C1}.

When assessing the other periods separately, we find the effect of β_1, the first lag of prices itself, to be persistent and significant in all periods. Surprisingly do we find the effect of temperatures β_2 only to be insignificant in the Full-Period, but significant in all separate periods. The coefficient fluctuates between being positive and negative and to be strongest in the Post-C19 period, although remains incremental. The coefficient for parameter β_3, the dummy variable of high relative temperatures, is negative among all periods but not significant in all periods. Considering coefficients for parameter β_3, high relative temperatures, we find to be strongly negative and significant in periods 2014, 2015, 2018, 2019. For coefficient of parameter β_4, low relative temperatures, we find the coefficient to be significant in periods 2016, 2017, 2018 and 2019. In these significant periods we find (strongly) positive results. The effect seems to be stronger for relative high temperatures. We however do not consider this enough robust evidence to accept H_{0C2}.

Considering the interaction terms, we find the coefficient of parameter β_5 to be slightly positive in all periods and significant in periods 2014, 2015 and 2019.

For the last coefficient, concerning parameter β_6, involving the dummy variable for low relative prices, we find a slightly negative coefficient in the Full-Period, but we find a positive coefficient in period 2018. The positive coefficient implies that during low relative prices, the slope of the regression between prices and temperature is steeper than normal. However, the low coefficients for β_5 & β_6 have to be noted which makes drawing conclusions from these coefficients difficult.

Lastly, is the C19-Period the least significant. This can be due to few observations, and the different behavior of prices during this period, which endorses the choice to make this the crisis-period.

4.2 Daily

4.2.1 ARIMA-Type models

When assessing daily prices, the dynamic of results changes significantly. On one hand there is 24-times less observations, as the daily prices are an average of the 24-hourly determined prices. Therefore, the 'daily price' is in fact not a real price at all but merely an average. Also do the external influences of exogenous regressors change which are short-term by nature, and short-spikes due to short periods of for example high windvelocities can be less well spotted. Advantages are that models are easier to handle and take shorter to compute and allow for easier longer-term forecasting.

From the ACF & PACF plots from the Pre-C19 period (Figure 8) we would expect an AR-lag of 5, 6 or 7 to fit best. We find in the Pre-C19 period that an AR(7) term performs best according to in-sample goodness-of-fit tests as displayed in Table 7, and an $MA(1)$ term to be optimal. However when combined we find that the combination of $AR(5)$ and $MA(1)$ only slightly underperforms in-sample the ARMA(7,1) model based on AIC and Log-Likelihood criteria. As data is stationary differencing is not necessary and an I(0) term should be optimal. Despite evidence of stationarity (see descriptive statistics) we found the same model with an I(1) component to have higher in-sample performance measures and information criteria. We find the ARMA(7,1) to be optimal for more extended models in 6.3 in both the Pre-C19 and Post-C19 period.

We again do not find evidence of our expectation that simple ARIMA-type models perform relatively worse during the C19-Period.

Table 7: In-sample goodness-of-fit tests by information criteria AIC & Log-Likelihood, daily frequency

	Pre-C19		**Post-C19**	
Model	**AIC**	**LogLik**	**AIC**	**LogLik**
ARIMA(1,0,0)	9.307	-4.652	322	-158
ARIMA(5,0,1)	9.124	-4.554	323	-153
ARIMA(5,1,1)	9.119	-4.553	319	-152
ARIMA(7,0,1)	9.125	-4.553	323	-152
ARIMA(7,1,1)	9.118	-4.551	325	-154
SARIMA(5,1,1)(1,0,0)[12]	9.121	-4.553	320	-152
SARIMA(5,0,1)(1,0,0)[12]	9.157	-4.570	324	-153
ARIMAX(5,1,1)	8.818	-4.400	312	-147
ARIMAX(7,0,1)	**8.824**	**-4.400**	318	-147
ARIMAX(7,1,1)	8.820	-4.399	316	-147
ARMAX(5,1)	8.823	-4.402	315	147
ARX(5,0,0)	8.843	-4.412	313	-147
ARMA(5,1)ARCH(2) 'norm'	8.923	-4.549	298	-152
ARMA(5,1)ARCH(2) 'std'	8.824	-4.419	266	-133
ARMA(5,1)ARCH(2) 'sstd'	9.252	-4.411	259	-132
ARMA(5,1)GARCH(1,1) 'norm'	9.006	-4.387	303	-152
ARMA(5,1)GARCH(1,1) 'std'	8.906	-4.337	266	-133
ARMA(5,1)GARCH(1,1) 'sstd'	9.338	-4.328	**252**	**-126**
RSGARCH	16.177	-8.078	458	-219

The best performing models have been highlighted in **Bold** which will be evaluated in Section 6.3.

4.2.2 ARIMA-Extensions

From the decomposed data in Figure 9 we observe seasonality is strongly present, and the data is subject to trending although it is not as strongly present as the seasonal component is. We chose not to detrend and deseasonalize in order to show which models can capture the characteristics of the raw data best and to set $m = 12$ because the annual seasonality is the clearest seasonal pattern. From the random component we can conclude that the prices behave spikey as expected (which provides potential for increased performance of GARCH-type models). We optimized by trial-and-error the ARIMA(5,0,1) model for the seasonal components (P,D,Q) and found this to be optimized in-sample at (1,0,0). The SARIMA-models barely improve the simpler ARIMA-type models, as they barely show a higher AIC and higher Log-Likelihood statistics than their simpler ARIMA-versions. There is no clear difference between the Pre-C19 period and the Post-C19 period visible from the SARIMA-type information criteria.

From the in-sample performance measures we find the RMSE of the SARIMA-type models to be equal or higher than their simpler ARIMA-versions, but there is a difference in the Post-C19 period, as we do observe improved RMSE and MAE measures. It appears that SARIMA-type models in the daily frequency perform better in the Post-C19 period than in the Pre-C19 period. These findings differ strongly from the intraday period, as we found the SARIMA-type models to show the relatively strong information criteria and the strongest performance metrics.

Adding exogenous variables, we find the ARIMAX-type models to show better information criteria than the simpler models, than the SARIMA-type models. Judging from the performance metrics (Table 8), we find a similar pattern. The ARIMAX-type models show in both the Pre-C19 and Post-C19 periods the strongest RMSE and MAE metrics. Adding the exogenous regressors certainly adds to the in-sample performance of models. Especially the windspeed variable has a negative influence onto prices (the negative relation of windspeeds onto prices will be extended in the OLS-analysis) as displayed in the coefficients in Table 10.

During the Post-C19 period, we find significant differences. First of all, do we find the SARIMA-models to perform relatively better based on in-sample criteria, although still performing less well than simpler ARIMA-type models. The ARIMAX-type models behave similar to the Pre-C19 period. Striking is the change in sign and stronger influence of the exogenous variable temperature. Caution is however required when analyzing the Post-C19 period due to the small-sample size of the Post-C19-period, and the seasonality of the period (the period only considers spring as it occurs in March, April and May while the Pre-C19 period considers all seasons in equal proportions. Judging from the performance, via the criteria RMSE and MAE as displayed in Table 8, we observe the RMSE & MAE to be in line with the conclusion from the information criteria and we judge model ARIMAX(7,0,1) to be the optimal model to use going forward.

Table 8: In-sample performance measures

	Pre-C19 period		Post-C19 period	
Model	**RMSE**	**MAE**	**RMSE**	**MAE**
ARIMA(1,0,0)	4,72	3,33	5,70	3,62
ARIMA(5,0,1)	4,44	3,15	5,17	3,46
ARIMA(5,1,1)	4,44	3,16	5,18	3,29
ARIMA(7,0,1)	4,43	3,15	4,86	3,43
ARIMA(7,1,1)	4,44	3,15	5,43	3,38
SARIMA(5,1,1)(1,0,0)[12]	4,44	3,16	5,11	3,25
SARIMA(5,0,1)(1,0,0)[12]	4,48	3,18	5,09	3,49
ARIMAX(5,1,1)	4,03	2,83	4,78	3,00
ARIMAX(7,0,1)	**4,02**	**2,82**	**4,54**	**2,93**
ARIMAX(7,1,1)	4,03	2,83	4,75	2,95
ARMAX(5,1)	4,03	2,83	4,58	2,97
ARX(5,0,0)	4,05	2,84	4,59	2,96
ARMA(5,1)ARCH(2) 'norm'	4,46	3,15	3,42	5,20
ARMA(5,1)ARCH(2) 'std'	4,46	3,14	3,03	6,01
ARMA(5,1)ARCH(2) 'sstd'	4,46	3,15	3,42	5,20
ARMA(5,1)GARCH(1,1) 'norm'	4,44	3,15	3,42	5,20
ARMA(5,1)GARCH(1,1) 'std'	**4,45**	**3,14**	**3,03**	**6,01**
ARMA(5,1)GARCH(1,1) 'sstd'	4,45	3,15	3,36	5,71

The best performing models have been highlighted in **Bold** which will be evaluated in Section 6.3.

4.2.3 ARIMA-GARCH models

For the daily data frequency the ARMA(5,1) model was deemed most appropriate for the further GARCH-analysis. First of all an analysis on the residuals will be performed of this ARMA(5,1) model prior to proceeding with the GARCH-analysis.

In Figure 11 the residuals of the model are plotted. The ARMA-model appears to have captured a lot of the dynamic of the model, as the ACF-plot of the Pre-C19 period has strongly improved over the ACF-plot of Figure 8. The residuals however do appear to show heteroskedasiticity and volatility clustering appears to be present. The ARMA-model can hardly be improved upon judging from the ACF and residuals plots of the Post-C19 period, although some volatility clustering appears to be taking place, especially during the large volatility- and price-shock in April 2020 the ARMA-model is far off. The Pre-C19's residual density plot nearly shows normality although far from perfect. The Post-C19 period appears to be left-skewed.

In Table 12 and Table 13 more formal tests of these observations are performed. The ARCH-LM test results show there is an ARCH-effect present in the residuals in both periods. The Ljung-Box test for autocorrelation in the residuals shows high p-values in both periods, leading to a failure of rejection of the null, and correlations among the residuals are random and behave like white-noise implying that the

ARIMA-type model captures a high amount of the models dynamic. The Jarque-Bera test indicates the residuals are not perfectly normally distributed. In order to account for this, we will use Student-t distribution and a skewed version of the Student-t distribution.

The information criteria provide insights that the Student-t distribution seems to be the best fit in the Pre-C19 period and the skewed Student-t distribution is the best fit for the Post-C19 period. Moreover, does the ARCH(2) achieve slightly better information criteria metrics in both periods.
Judging from in-sample performance measures that little to no improvement is observable in the Pre-C19 period. This is a striking finding given that the input of the mean model is the best-fit ARIMA-type model. The ARMA-GARCH models in the C19-period do however improve strongly compared to the simpler version of the ARMA(5,1) model.
Surprising are the high values for the MAE for the ARMA-GARCH-type models in the Post-C19 period. This can be likely due to the high spike in the Post-C19 period (which can be observed from Figure 12). The ARMA-GARCH model is namely the best fitting model based on the RMSE overall in the Post-C19 period, while the best fitting models in the Pre-C19 period actually are ARMAX-type models. The best-fit model is the ARMA(5,1)GARCH(1,1)'std' as it has the lowest AIC. The ARMA-GARCH-type models do perform relatively better in the Post-C19 period compared to the pre-C19 period in the daily data-frequency.
From the weighted ARCH-LM test as displayed in Table 16, we reject the null-hypothesis in each period, for each GARCH-type model and for each lag of 3, 5 and 7 and we can conclude that there are no ARCH-effects remaining in the data and the residuals can be considered close to White noise in each tested model for each period. For the Post-C19 period we found the ARMA(5,1)GARCH(1,1) with a Student-t distribution to be best fitting and performing and will therefore be evaluated in Section 4.3.

4.2.4 GARCH Regime-Switching models

We find the regime switching GARCH models to underperform so strongly that we disregard them for the analysis. See Table 17 for the results.

4.3. Model Evaluation

4.3.1 Intraday

In this section we will evaluate the estimated models that proved to be best performing. From the intraday analysis, we found the SARIMA(24,3)(3,1,1)[24] to be best fitting and performing. The model especially outperformed during the Pre-C19 period. The model's coefficients are summarized in Table 9. The table shows many insignificant AR(p) and MA(q) to be insignificant. However, the SAR factors are all highly significant, except for the SAR3 coefficient. This indeed indicates that the SARIMA-type model fits the data well.

Table 9: Results from the SARIMA(24,3)(3,1,1)[24] model

Parameter	Pre-C19			Post-C19		
AR(1)	0,4712	0,4505	0,2956	0,3589	0,3351	0,2841
AR(2)	0,5419	0,1314	***	0,6360	0,1285	***
AR(3)	0,0835	0,2292	0,7156	0,1083	0,1911	0,5708
AR(4)	-0,2218	0,1238	*	-0,2906	0,1539	*
AR(5)	0,0052	0,0148	0,7250	0,0475	0,0343	0,1666
AR(6)	0,0034	0,0136	0,8028	0,0089	0,0242	0,7133
AR(7)	-0,0012	0,0113	0,9141	0,0260	0,0244	***
AR(8)	0,0144	0,0104	0,1654	-0,0221	0,0277	0,4239
AR(9)	-0,0100	0,0127	0,4341	-0,0518	0,0241	**
AR(10)	-0,0057	0,0086	0,5050	0,0734	0,0313	**
AR(11)	-0,0062	0,0081	0,4406	0,0483	0,0296	0,1030
AR(12)	0,0168	0,0082	**	-0,0419	0,0352	0,2343
AR(13)	0,0140	0,0096	0,1446	0,0284	0,0251	0,2585
AR(14)	-0,0066	0,0129	0,6098	-0,0130	0,0259	0,6155
AR(15)	-0,0039	0,0086	0,6546	-0,0281	0,0252	0,2644
AR(16)	0,0000	0,0080	0,9989	-0,0682	0,0260	***
AR(17)	0,0088	0,0081	0,2787	0,0602	0,0374	0,1069
AR(18)	0,0034	0,0091	0,7121	0,0236	0,0253	0,3519
AR(19)	0,0017	0,0093	0,8574	0,0007	0,0276	0,9792
AR(20)	-0,0064	0,0087	0,4602	-0,0075	0,0256	0,7703
AR(21)	0,0083	0,0079	0,2974	-0,0096	0,0247	0,6967
AR(22)	0,0127	0,0083	0,1248	-0,0045	0,0254	0,8599
AR(23)	0,0500	0,0109	***	0,1105	0,0238	***
AR(24)	-0,0503	0,0268	*	-0,0963	0,0298	***
MA(1)	0,3113	0,4504	0,4894	0,6816	0,3349	**
MA(2)	-0,3033	0,4467	0,4971	-0,1272	0,4200	0,7621
MA(3)	-0,3293	0,1634	**	-0,3275	0,2052	0,1105
SAR(1)	0,1270	0,0109	***	0,2498	0,0337	***
SAR(2)	0,0298	0,0075	***	-0,0498	0,0224	**
SAR(3)	0,0166	0,0073	**	0,0146	0,0210	0,4852
SMA(1)	-0,9065	0,0048	***	-0,9237	0,0122	***

*, **, *** correspond to significance at 10%, 5% and 1%

4.3.2 Daily

In this section we will evaluate the estimated models that proved to be best performing. From the daily analysis, we found the ARMAX(7,0,1) model to perform and fit best. For the Post-C19 period we found the ARMA(5,1)GARCH(1,1) with a Student-t distribution to be best fitting and performing.

In Table 10 the estimated coefficients from the ARMAX(7,0,1) model are shown. We observe the first AR lags to be significant. The first lag has a strongly positive influence on current prices. At the second AR-lag we observe the sign to change in the Pre-C19 period (in the Post-C19 period the $AR(2)$ lag is insignificant. The $MA(1)$-lag is furthermore significant and strongly negative in the Pre-C19 period. In the Post-C19 period the MA(1) lag is positive but insignificant. We furthermore observe the influence of both wind and temperature to be significant in the Pre-C19 period, but the effect of wind is overall stronger.

Table 10: Estimated ARIMAX(7,0,1) model

	Pre-C19 period			Post-C19 period		
Parameter	**Coef**	**Std**	**Sign**	**Coef**	**Std**	**Sign**
Intercept	48,7899	3,0559	***	23,8140	3,6326	***
AR(1)	1,5267	0,0526	***	0,3768	0,8496	
AR(2)	-0,5611	0,0539	***	-0,1215	0,5019	
AR(3)	0,0743	0,0485		-0,0582	0,2642	
AR(4)	0,0003	0,0487		-0,0425	0,1458	
AR(5)	0,0057	0,0487		0,1162	0,1671	
AR(6)	-0,0258	0,0466		0,0719	0,1917	
AR(7)	-0,0228	0,0288		-0,1317	0,1386	
MA(1)	-0,9145	0,0458	***	0,2014	0,8524	
Temperature	-0,0130	0,0045	**	0,0185	0,0268	
Wind	-0,1478	0,0086	***	-0,1775	0,0494	***

Where 'Coef' stands indicates coefficient, 'Std' stands for standard deviation and 'Sign' for significance

The model for the daily period's best performing model are found in Table 11. There is a clear GARCH-effect in the Pre-C19 period. The conditional volatility relies heavily on the β_1 coefficient in both periods, implying current volatility relies strongly on past volatility. The $\propto_1$ coefficient is only significant in the Pre-C19 period, which implies the dependency of the conditional volatility on its past errors. The Omega ω coefficient is merely significant in the Pre-C19 period.

All AR and MA coefficients except $AR(2)$ are and statistically significant for the C19-period, and not surprisingly does the $AR(1)$ coefficient have the most significant positive influence onto current mean in time t. The coefficient β_1 is statistically significant, $\propto_1$ and ω are insignificant. This implies we can only conclude time-varying volatility depends strongly on its own past residuals at a rather low value in the Post-C19 period. All coefficients meet the stability condition of $\propto_1 + \beta_1 < 1$. As the coefficients sum up closely to one, this implies that volatility shocks are quite persistent; high volatility tends to be followed by large changes and low volatility tends to be followed by small changes in volatility. The Pre-C19 period more so than the Post-C19 period.

Table 11: Results summary from the ARMA(5,1)(GARCH(1,1) model with Student-t distribution

	Pre-C19 period			**Post-C19 period**		
	ARMA(5,1)GARCH(1,1)std			*Robust Standard Errors*		
Parameter	**Est**	**Std**	**Sign**	**Est**	**Std**	**Sign**
μ	41,51	2,95	***	22,15	0,42	***
AR(1)	1,20	0,08	***	0,50	0,15	***
AR(2)	-0,28	0,06	***	0,04	0,06	
AR(3)	-0,04	0,04		-0,12	0,05	***
AR(4)	0,05	0,04		0,14	0,05	***
AR(5)	0,05	0,03		0,22	0,04	***
MA(1)	-0,62	0,08	***	-0,48	0,13	***
ω	11,08	0,34	***	16,50	19,77	
$\propto_1$	0,17	0,03	***	0,00	0,14	
β_1	0,78	0,04	***	0,82	0,15	***
Shape	51,63	0,60	***	21,00	0,01	***

, **, * correspond to significance at 10%, 5% and 1%. 'Est' stands for estimated coefficient. 'Std' stands for standard deviation and 'sign' stands for significance. Robust standard errors were used given heteroskedasticity.*

When analyzing the residuals of the ARMA(5,1)GARCH(1,1) model as summarized in Table 16, we observe from the Pearson Goodness of Fit test that the normal distribution is an imperfect fit, given the low p-values. The Student-t has a good fit. In the Post-C19 period the normal fit is a good fit according to the Pearson Goodness of Fit test. From the Ljung-Box test on the standardized squared residuals we conclude that we fail to reject the null hypothesis of autocorrelation among the squared residuals. The residuals can be considered close to White noise indicating a strong fitting model in the Pre-C19 period. In the Post-C19 period the null-hypothesis is rejected as the p-values are all significant at the 5% level for the different lags measured. Hence, there room for improvement considering the Post-C19 GARCH-fit. Given the high p-values the weighted ARCH-LM test provides, we conclude that there is no more ARCH-effect present in neither the Pre-C19 period nor the Post-C19 period.

5. Conclusions

The scope of this paper is to model the energy market with many of its characteristics and changing dynamics, in order to assess which models worked best during normal-periods and a crisis-period in both daily and intraday frequency. We modeled the energy market using ARIMA, SARIMA, ARIMAX, ARIMA-GARCH and Markov-Switching-GARCH type models in both periods in intraday and daily data-frequency. This paper puts new perspective on the value of especially windspeeds and to a lesser degree temperatures onto Dutch Day-Ahead prices, and finds strong modeling performance of GARCH-type models during the crisis period that Covid-19 provided, both in daily- and intraday models.

The general research question was: *"Which model-types model the dynamics of the Dutch Day-Ahead market best, and how does the Pre-C19 period compare to the Post-C19 period in terms of these criteria?"*

Summarizing, the best in-sample performance differs among periods. We find the SARIMA models are the best performing models in the intraday data frequency, especially in the Pre-C19 period. ARIMAX-type models perform less well than expected in the intraday-frequency. Concerning the daily data frequency, we find the ARIMAX model to best model the market dynamics. As expected, GARCH-type models do not perform best in the intraday frequency.

In the daily data-frequency, we find SARIMA-type models to perform relatively better during Post-C19 period, although this model-type underperforms in general in the daily data-frequency. ARIMAX-type models perform better in the daily data-frequency. The ARMA-GARCH-type models do perform relatively better in the Post-C19 period compared to the pre-C19 period in the daily data-frequency.
A surprising finding that the GARCH models do not outperform ARMAX models in the Pre-C19 daily period, although this does support our expectation that GARCH dynamics improve modelling performance in a crisis-period.

From the OLS-extension, we find a negative relation between windvelocities and heights of prices and the magnitude increases in more extreme quantiles. Despite the appearance of the relation becoming stronger over time, as the renewables share grows in the energy mix, we judge this pattern as not clear enough to conclude the effect has increased in the last 7 years. We furthermore find extreme prices go hand-in-hand with relatively extreme temperatures. We moreover find that during high relative temperatures the intercept of Prices is strongly more negative, and the slope of the regression is slightly more positive, where the effect is opposite (and a bit less severe) for low relative prices.

In the intraday and in the daily frequency, we do not necessarily find simple ARIMA models to perform less well in the Post-C19 period compared to the Pre-C19 period, given this expectation resulting from the increased volatility-clustering in this period. A further counterintuitive finding is that we do not find stronger performance of the GARCH-type models in the intraday data frequency during the Post-C19 period.

When comparing both periods, and the performance of GARCH-models, residuals can be considered close to White noise in each tested ARMA-GARCH model in the daily-frequency, and only in one model in the hourly period. Thereby we can conclude in our data sample the GARCH-type model performs better in the daily data frequency.

6 Implications & Limitations

6.1 Economic & Academic Implications

As the weather will likely become more often extreme, and the share of renewables will grow as strongly as projected, grid-operators and energy companies need to be ready to match supply and demand to prevent grid-problems such as in Texas in February 2021. Modeling prices and forecasting well will be of increased importance for agents in the power market.

An increased share in renewable energy may also have implications for electricity prices, as we find during high windvelocity periods that prices are lower than during normal and low windvelocities. This may be indicative of a situation where wind generated electricity dominates the market for a short time, although one needs to be careful of other possibilities of this causation. The non-storability of electricity also likely has a high influence on these lower prices, since generators need to 'dump' their electricity.

We can not conclude, despite some but not robust enough evidence, that the relation between extreme windspeeds and extreme temperatures is becoming stronger within our sample. We are however limited to the short-time in which renewably generated electricity has come to the grid. It can be reasonably expected that this source-type will continue to grow in the near future, and that the effect will (more robustly) stronger. This would have large implications for the electricity markets, given the near-zero marginal costs these sources' operators face, will likely cause the power price to go down in the long-term. This would be a great situation form a consumer point-of-view, but less optimal for operators and new investments into this sector.

When modeling the electricity market, one needs to carefully examine the drivers of the market, and likely use a combination of complex models in order to achieve the best performing models and most accurate forecasts.

6.2 Limitations & Future Research

A limitation of this research is that the crisis period, especially for the daily sample, has a relatively limited amount of data. This short crisis-period is furthermore only during a few months in the spring-season, which is also a limitation of the comparison method. These limitations will however often be the case when assessing a period of crisis, and is in line with previous comparison research papers in the literature.

Furthermore, it requires to be mentioned that the obtained weather data (temperature and windvelocity) are stored 'homogeneously' which implies that it might have occurred that the way data has been measured is not consequent, which should be noted. Furthermore, is this data from De Bilt, while the price data is nationwide and the correlation may sometimes be off, but it is considered the best proxy.

The metrics used for performance, are limited to two metrics. It can occur that some other metrics would be more suited for one type of model, as metrics tend to be favorable to certain characteristics. With RMSE we tried to use the most neutral metric for all models but RMSE is known to strongly punish when predictions deviate strongly from the actual observations. For example during the negative shock resulting from the Covid-19 crisis, some models may be far off with predicted values versus observed values. This makes the crisis-period more prone by nature, and give worse RMSE values unless modeled perfectly.

The purpose of this study was to study the dynamics of the power market, for future research it may be interesting to combine some of the better performing models as presented in this study in order to achieve higher in-sample (and out-of-sample) forecasting accuracy. Furthermore, is an interesting topic of research the characteristics of the imbalance market. This market shows many more spikes and extreme behaviors which makes modeling a tough though interesting challenge, and the market has not been reviewed that often in the literature. Partly because this market is more difficult to forecast given its strong spikey behavior. When one is able to forecast the Day-Ahead market and imbalance market well, one can trade on both markets which may balance out the markets volatility.

When one purely wants to assess the implication of a higher share of renewable energy onto the grid, one may assess a different country such as Germany in which the renewable share is more significant. This poses an interesting future research topic.

Lastly, looking backward, I experienced some technical difficulties which can be considered limitations. By acknowledging them here, future researchers of similar topics may gain an advantage by prior knowledge of them. First of all was the data for the imbalance market corrupted. This made me switch to the in the literature more widely covered and in general better forecastable Day-Ahead market.

Furthermore, was I forced to shrink my dataset concerning the intraday price-analysis, for computational simplicity by two years from 2014 as a starting point to 2016 as a starting point. Lastly, did I try to assess the models on a more sophisticated basis, with the Model Confidence Set as described by Hansen et al. (2011), but unfortunately the execution proved too difficult working with all the different packages in the program R.

Appendix

Figures

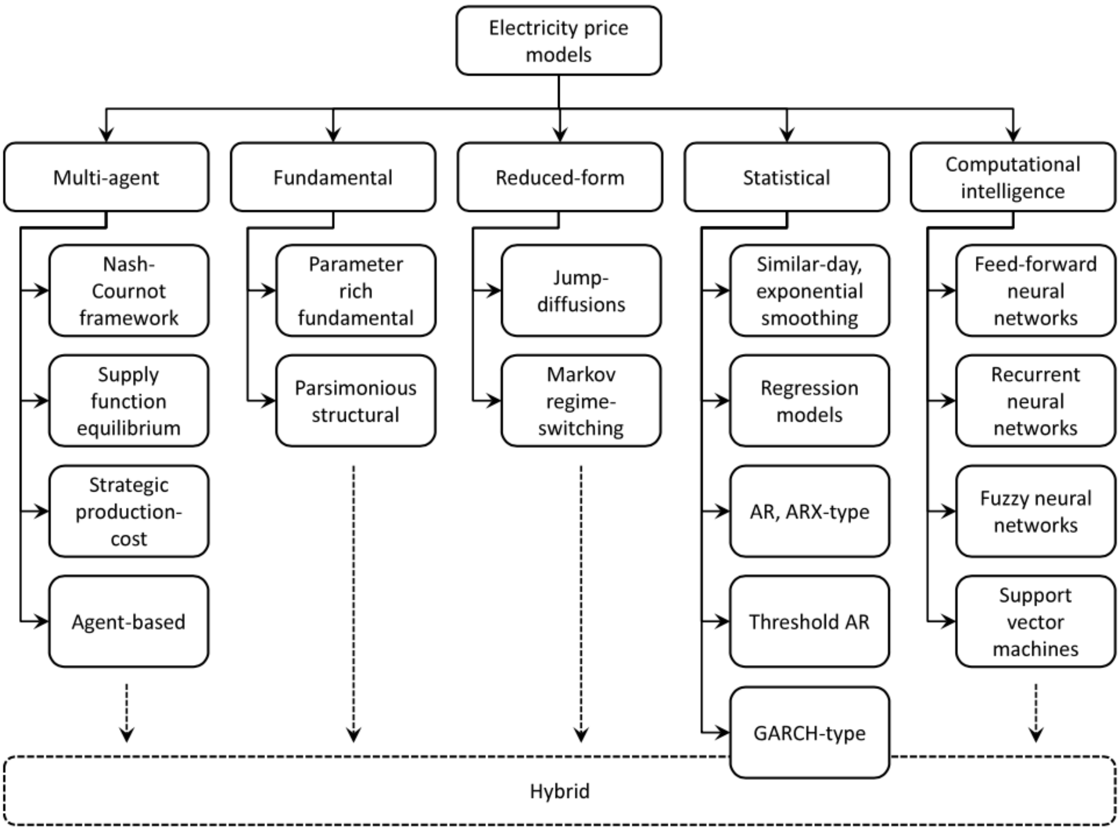

Figure 3: Adopted from Weron (2014). A taxonomy of electricity spot price modeling approaches. Focus in this thesis is on Reduced-form, Statistical and Fundamental models.

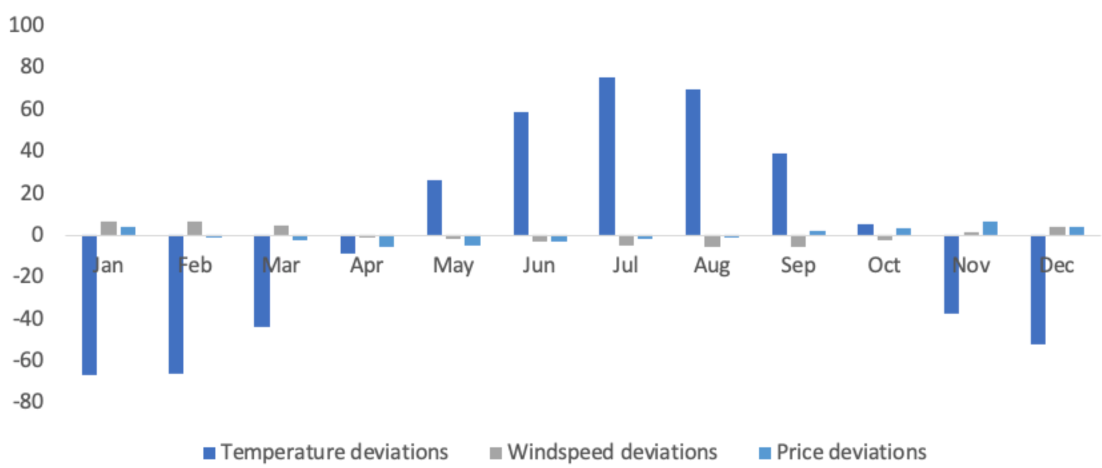

Figure 4: Deviations from the average for temperature, windspeeds and Day-Ahead prices in high-frequency (intraday) data.

Based on the findings from the average price deviations, we set the dummy for peak-hours equal to 1 during hours 8, 9, 10, 11, 12, 13, 18, 19, 20, 21.

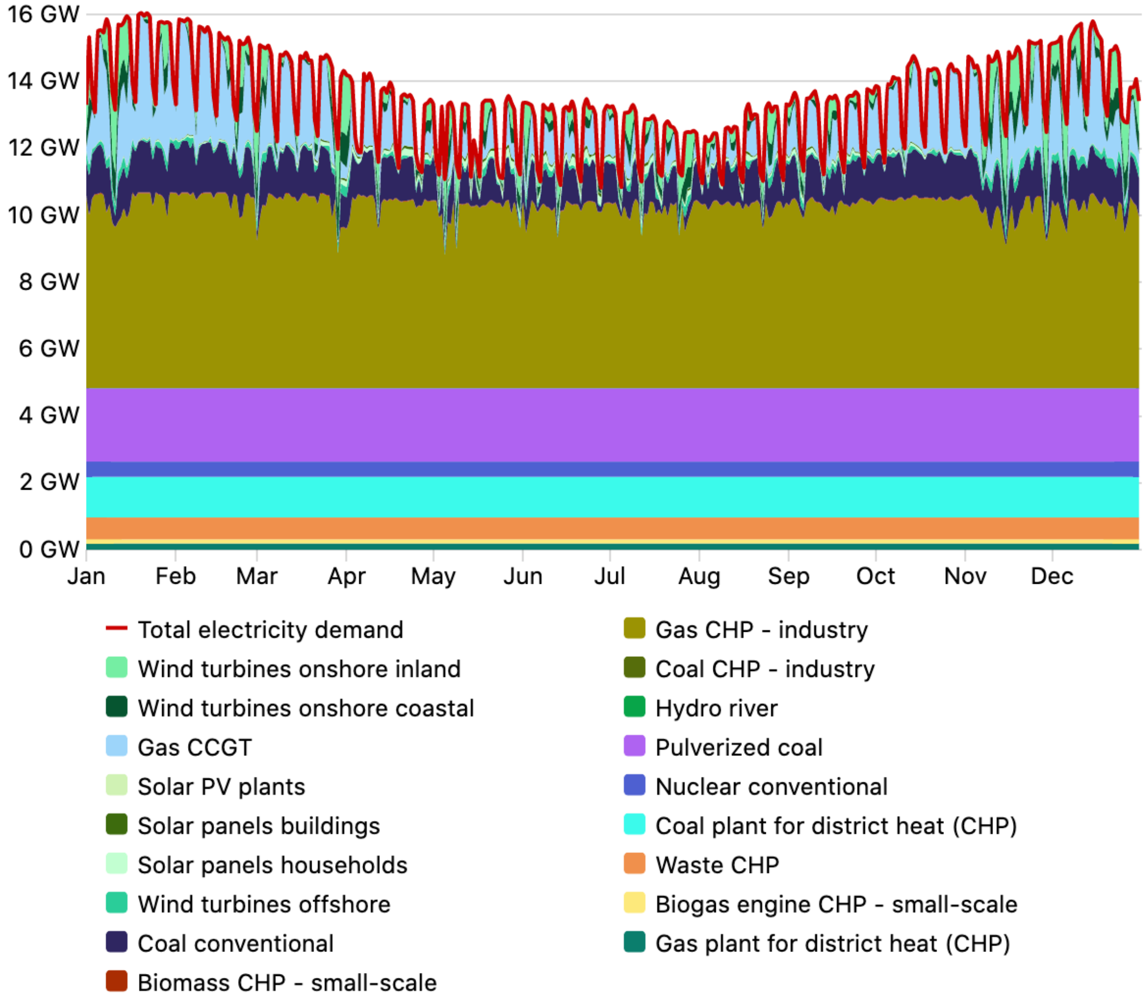

Figure 5: This chart shows the hourly production of electricity and gives a visual clue of the electricity generation mix of the Netherlands.

Source: Energy Transition Model (pro)

We observe nuclear, coal and gas to a certain extend to have a rather stable power supply. Moreover, do we observe that the variability form total electricity demand is captured by more flexible sources, especially wind and solar sources in all sorts of forms and to a lesser extend coal ('conventional') and gas ('CCGT').

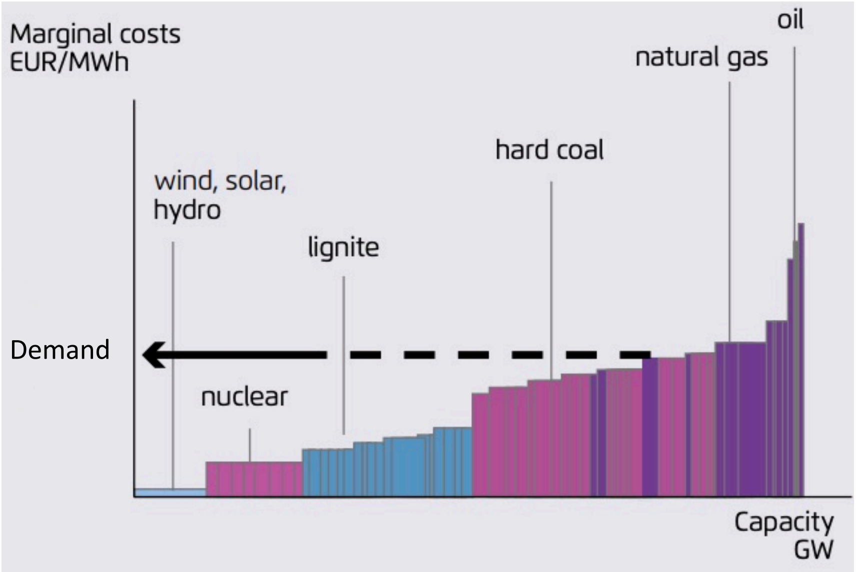

Figure 6: The merit-order curve.

Source: European Commission

In this figure the merit-order curve is depicted of European electricity production. On the y-axis the marginal costs in EUR/MWh are observed. On the x-axis the capacity in Gigawatt (GW) is shown. Electricity generating sources are sorted from low (left) to high (right). The demand curve is depicted in black. Renewables have much lower marginal costs than for example coal, gas and oil. The figure not only shows which generating sources have the highest marginal costs, but also that when prices (demand) are low, some sources are loss-making.

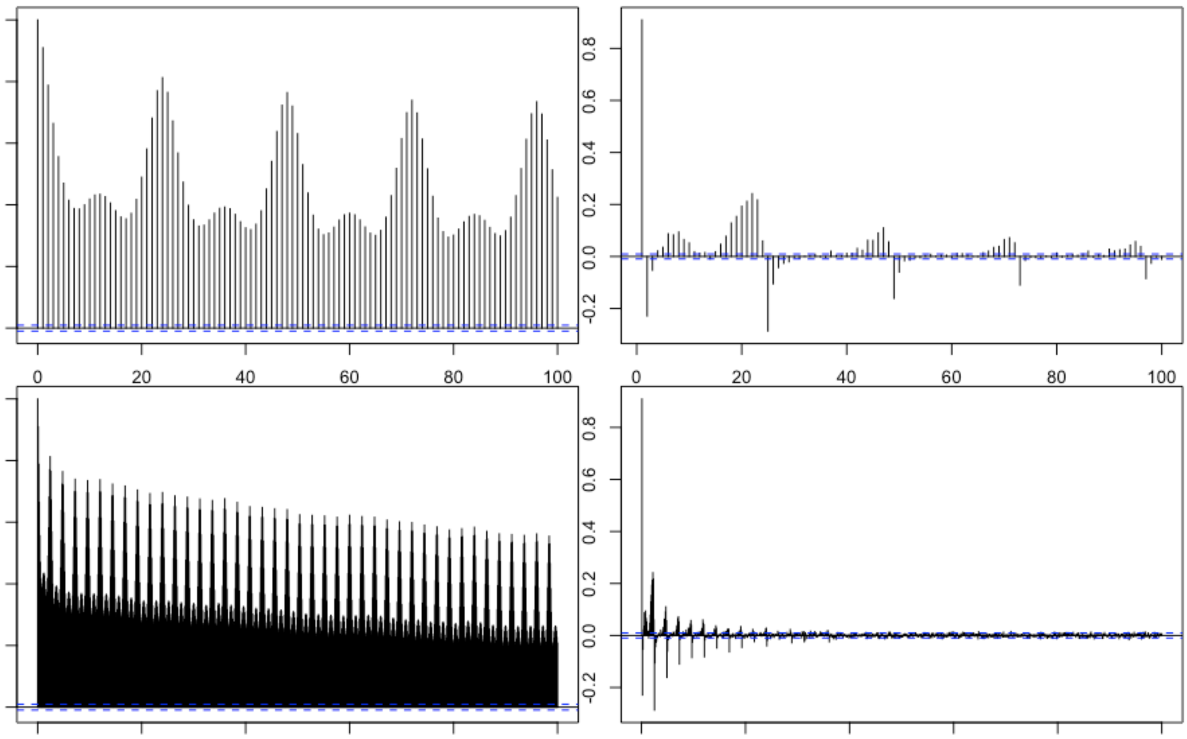

Figure 7: ACF and partial ACF (PACF) plots are graphically represented from the intraday data over the full sample 2014-2020. On the left-hand side the ACF plots are shown, and on the right-hand side the PACF plot is shown. Both show a maximum lag up to 100.

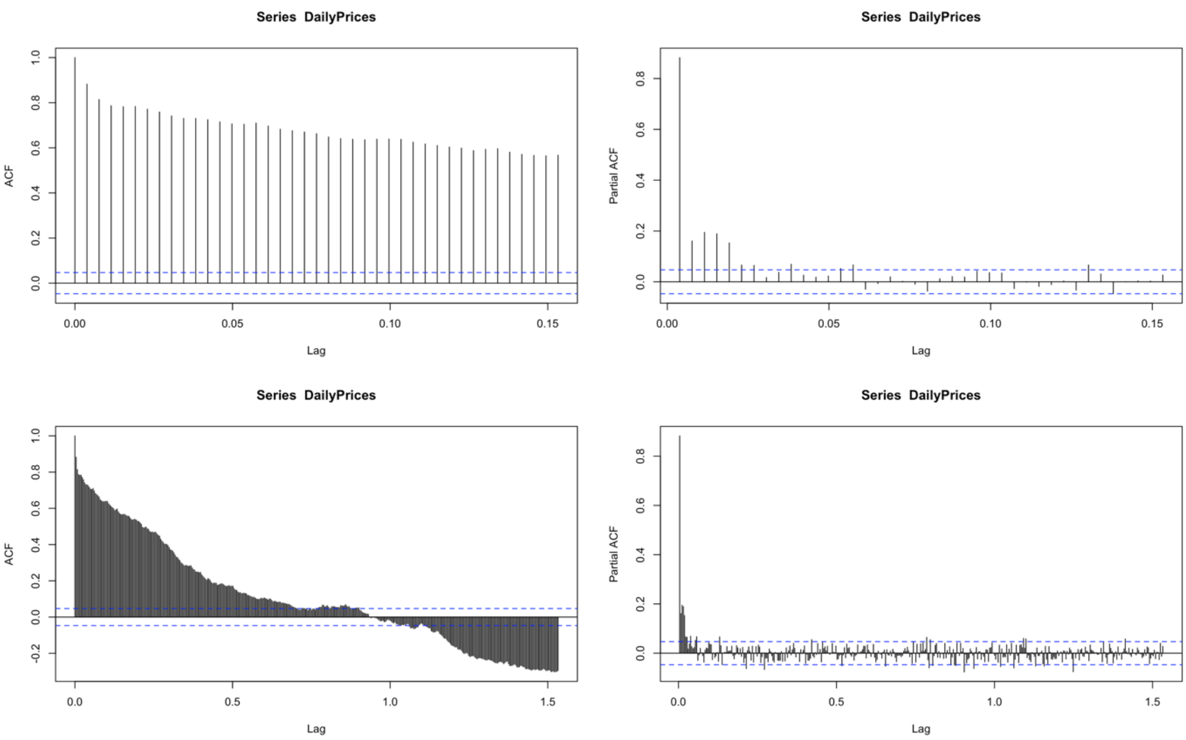

Figure 8: ACF and partial ACF (PACF) plots are graphically represented for the daily dataset containing Day-Ahead prices of the pre-crisis period 2014-2019. On the left-hand side the ACF plots are shown, and on the right-hand side the PACF plot is shown. The two upper plots have 40 lags, while the lower figures show 400 lags.

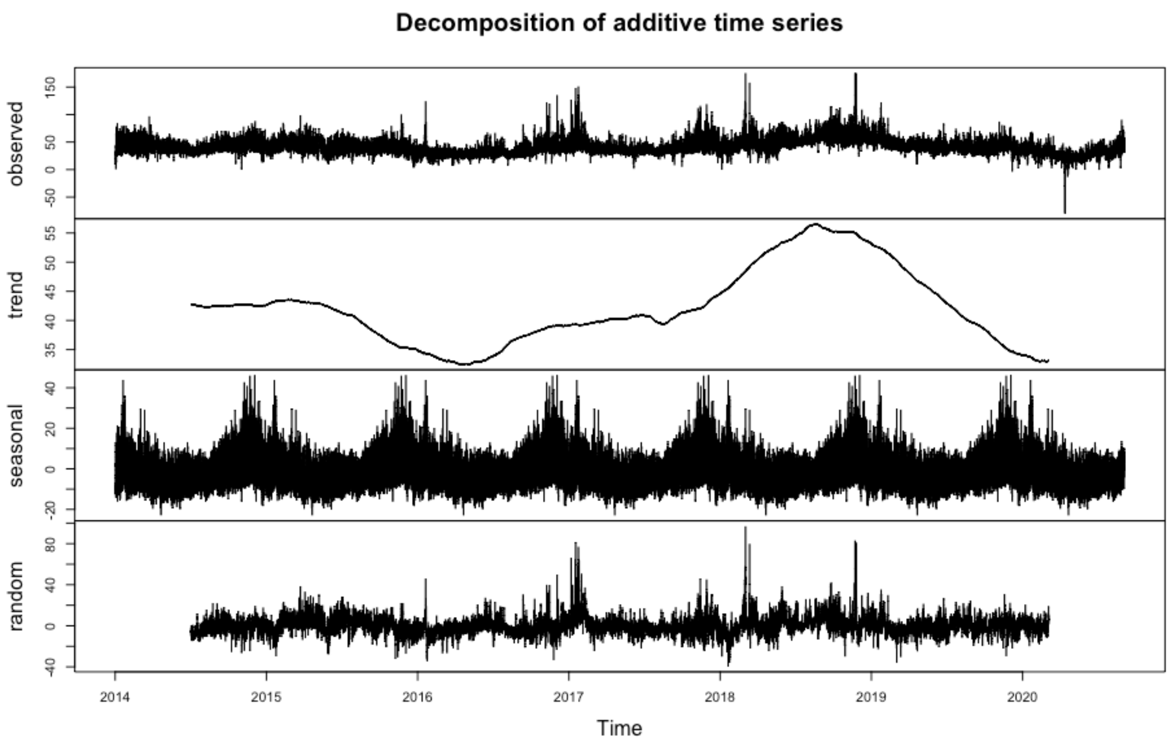

Figure 9: Decomposition of the additive intraday Dutch APX price (2014-2020, full sample).

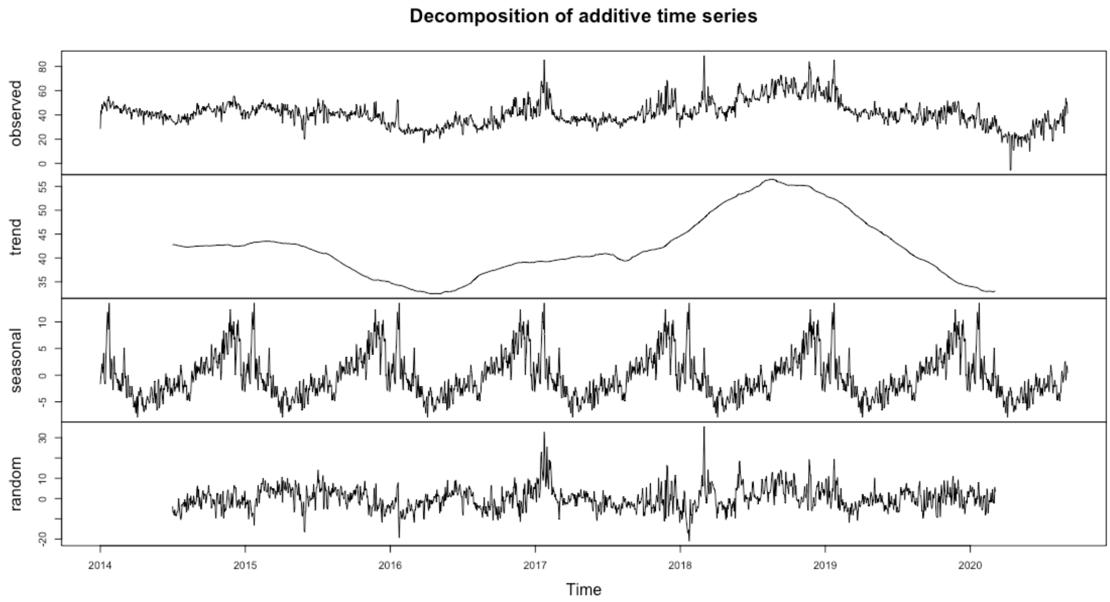

Figure 10: Decomposition of the (multiplicative) Dutch daily average APX price (2014-2020, full sample)

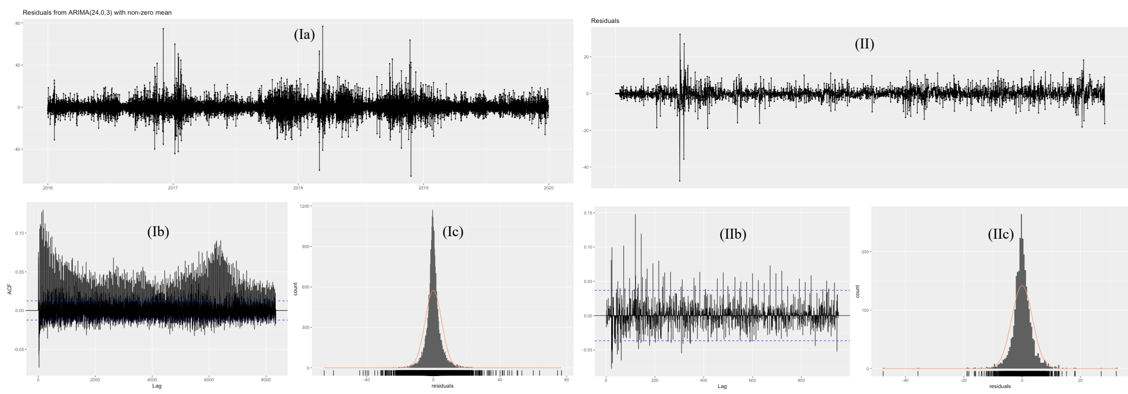

Figure 11: Residuals of the intraday ARIMA(24,0,3) model in the Pre-C19(I) and Post-C19(II) sample.
Table 'a' is the plotted time-series of the models residual, 'b' plots the models autocorrelation function ACF) and 'b' refers to the density plot of the residuals.

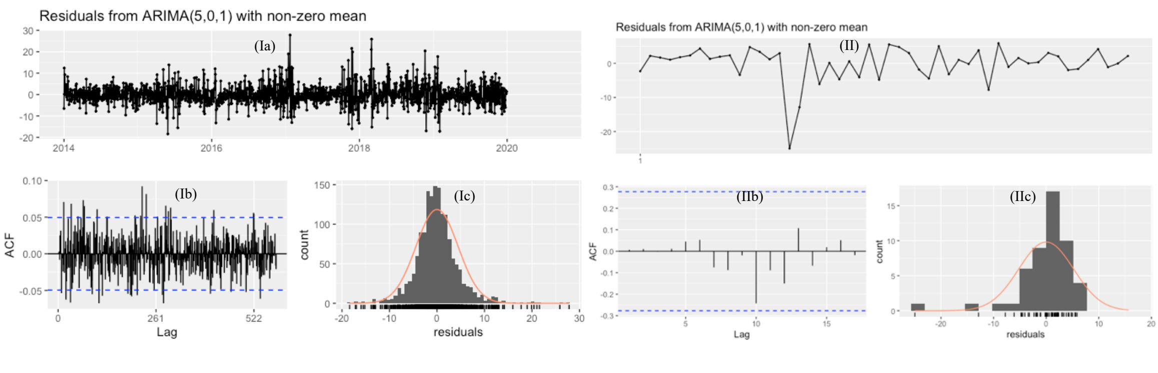

Figure 12: Residuals of the Daily ARIMA(5,0,1) model in the Pre-C19 (I) Post-C19 (II) sample.

Table 'a' is the plotted time-series of the models residual, 'b' plots the models autocorrelation function ACF) and 'b' refers to the density plot of the residuals.

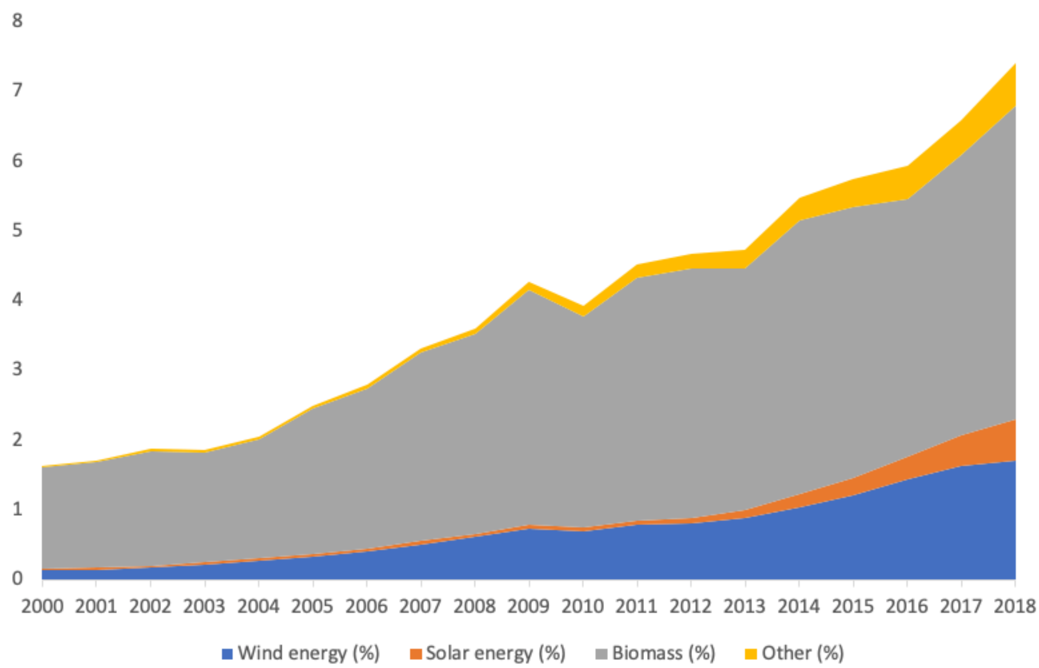

Figure 13: The increasing share of renewable generated sources is increasing strongly (period 2000-2018).
Source: CBS Statline

Tables

Table 12: Analysis of the residuals of the best fit ARIMA-type models

				Statistic		Critical-value
Test	H0	Period	Data	Statistic (Chi-squared / X-squared)	Lags	p-value
ARCH-LM	$\propto_0=\propto_1=\cdots=\propto_n$	Intraday, Pre-C19	Residuals ARMA(24,3)	548.01	22	<0,01
		Intraday, Post-C19	Residuals ARMA(24,3)	44,88	22	<0,01
		Daily, Pre-C19	ARMA(5,1)	165,81	22	<0,01
		Daily, Post-C19	ARMA(5,1)	29,91	16	0,0185
Ljung-Box	$\rho_1=\rho_2=\cdots=\rho_k$	Intraday, Pre-C19	Residuals ARMA(24,3)	548.01	22	<0,01
		Intraday, Post-C19	Residuals ARMA(24,3)	44,88	22	<0,01
		Daily, Pre-C19	ARMA(5,1)	0,00	1	0,99
		Daily, Post-C19	ARMA(5,1)	0,00	1	0,96
Jarque-Bera	$S = EK = 0$	Intraday, Pre-C19	Residuals ARMA(24,3)	315.950	2	<0,01
		Intraday, Post-C19	Residuals ARMA(24,3)	52.268	2	<0,01
		Daily, Pre-C19	ARMA(5,1)	1151	2	<0,01
		Daily, Post-C19	ARMA(5,1)	221,73	2	<0,01

Intraday

Table 13: Test statistics residuals intraday ARMA-GARCH models

Model Type	**ARMA(24,1)GARCH(1,1)norm**		**ARMA(24,1)GARCH(1,1)std**		**ARMA(24,1)GARCH(1,1)sstd**		**ARMA(24,1)GARCH(0,2)norm**		**ARMA(24,1)GARCH(0,2)std**		**ARMA(24,1)GARCH(0,2)sstd**	
Model Nr.	**I**		**II**		**III**		**IV**		**V**		**VI**	
Pre-C19 Period Hourly	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
Pearson Goodness of Fit (adjusted)												
Group 20	2249	0	250,1	2,93E-39	219,6	4,17E-33	5897	0	297,5	6,41E-49	274,9	2,74E-44
Group 30	2377	0	277,2	2,48E-39	239,9	4,68E-32	6059	0	297,9	2,09E-43	276,8	2,99E-39
Group 40	2512	0	296,3	2,75E-38	246,3	6,64E-29	6224	0	316,6	3,53E-42	299,2	7,71E-39

Group 50	2585	0	321,8	8,88E-39	275	3,29E-30	6373	0	317,6	5,32E-38	303,9	1,80E-35
Ljung-Box on standardized residuals												
Model Nr.	**I**		**II**		**III**		**IV**		**V**		**VI**	
Lag[1]	14,84	0,0001168	9	0,002188	6	0,01624	5	0,02177	169,3	0	166	0
Lag[5]	1374,29	0	953	0	961	0	1.146	0	991,3	0	990,5	0
Lag[9]	1915,49	0	1.634	0	1.660	0	1.656	0	1718,1	0	1723	0
d,o,f=2												
Ljung-Box on standardized squared residuals												
Model Nr.	**I**		**II**		**III**		**IV**		**V**		**VI**	
Lag[1]	222	0	245,5	0	237,9	0	2273	0	2615	0	2607	0
Lag[5]	236,9	0	262,8	0	255,1	0	2804	0	3187	0	3178	0
Lag[9]	269,9	0	296,1	0	289,4	0	2955	0	3326	0	3317	0
d,o,f=2												
Weighted ARCH-LM test												
Model Nr.	**I**		**II**		**III**		**IV**		**V**		**VI**	
Lag[3]	2.736	9,81E+01	3.233	7,22E+01	3.522	6,06E+01	283,5	0	311,5	0	311,3	0
Lag[5]	27.427	2,83E-04	29.128	1,05E-04	30.230	5,49E-05	368,3	0	380,9	0	380,9	0
Lag[7]	51.890	2,22E-10	53.434	8,48E-11	55.285	2,68E-11	396,2	0	398,4	0	398,3	0

Post-C19 Period Hourly	**ARMA(24,1)GARCH(1,1)norm**		**ARMA(24,1)GARCH(1,1)std**		**ARMA(24,1)GARCH(1,1)sstd**		**ARMA(24,1)GARCH(0,2)norm**		**ARMA(24,1)GARCH(0,2)std**		**ARMA(24,1)GARCH(0,2)sstd**	
Model Nr.	**I**		**II**		**III**		**IV**		**V**		**VI**	
	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
Pearson Goodness of Fit (adjusted)												
Group 20	20	133,7	2,92E-16	0,07582	29,29	0,0615	431,3	1,30E-76	19,65	0,4156	17,96	0,5251
Group 30	30	148,6	5,24E-15	0,13053	32,37	0,304	464,8	4,64E-77	37,37	0,137	37,75	0,1281
Group 40	40	175	3,79E-16	0,26677	45,64	0,2156	487,1	9,31E-76	40,26	0,4144	37,93	0,5185
Group 50	50	188,4	3,16E-15	0,24442	57,45	0,1906	515,7	4,22E-76	54,93	0,2601	55,46	0,2444
Ljung-Box on standardized residuals												
Model Nr. I			**II**		**III**		**IV**		**V**		**VI**	
Lag[1]	7	0,00777	1.754	0,1854	1,83	0,1761	0,2605	0,6098	1	0,2735	1	0,2691
Lag[5]	164	0	137	0	137,75	0	118	0	160	0	160	0
Lag[9]	269	0	239	0	240,24	0	189	0	234	0	233	0
d,o,f=2												
Ljung-Box on standardized squared residuals												
Model Nr.	**I**		**II**		**III**		**IV**		**V**		**VI**	
Lag[1]	2	0,1958	2	0,1853	2	0,1617	12,13	0,000497	16,91	3,92E-02	17	3,73E-02
Lag[5]	5	0,1289	4	0,2199	5	0,1925	421,45	0	523,66	0,00E+00	524	0,00E+00
Lag[9]	8	0,1202	7	0,2232	7	0,1975	490,9	0	599,96	0,00E+00	600,4	0,00E+00
d,o,f=2												
Weighted ARCH-LM test												

Lag[3]	2	0,19306	0,3407	0,5594	0,3706	0,5427	1	2,32E+00	10,31	1,32E+00	10,33	1,31E+00
Lag[5]	7	0,03709	46	0,1269	46	0,1259	45	9,25E-09	41,23	8,50E-08	41,33	8,03E-08
Lag[7]	8	0,0635	56	0,1705	56	0,1679	57	9,33E-12	51,1	3,63E-10	51,21	3,40E-10

Table 14: Markov/Regime Switching GARCH model estimates and (state) probabilities

	Coef	**Sign**	**Coef**	**Sign**
α_{0_1}	85	***	36	**
α_{1_1}	0,9997	***	0,9999	***
β_1	0,0002	**	0	***
α_{0_2}	852.129	**	36	***
α_{1_2}	0,9997	***	0,9999	***
β_2	0,0002	***	0	***
P_{1_1}	0,995	***	0,6613	***
P_{2_1}	0,4523	***	0	**

Where 'Coef' stands indicates coefficient and 'Sign' for significance.

State 1 prob	0,989	State 1 prob	0
State 2 prob	0,011	State 2 prob	1

	t+1|k=1	t+1|k=2		t+1|k=1	t+1|k=2
t|k=1	0,995	0,005	t|k=1	0,6613	0,3387
t|k=2	0,4523	0,5477	t|k=2	0	1,0000

Daily

Table 15: Estimated coefficients Daily models

PreC19	**Mean**	**AR1**	**AR2**	**AR3**	**AR4**	**AR5**	**AR6**	**AR7**	**MA1**	**SAR1**	**SAR2**	**XTemp**	**XWind**
ARIMA(1,0,0)	42,7	0,87											
ARIMA(5,0,1)	42,5	1,43	-0,51	0,04	0,04	-0,02			-0,82				
ARIMA(5,1,1)		0,55	-0,03	0,02	0,03	0,06							
ARIMA(7,0,1)	42,5	1,53	-0,56	0,04	0,02	0,02	-0,02	-0,02	-0,91				
ARIMA(7,1,1)		0,57	-0,02	0,03	0,04	0,06	0,01	0,04					
SARIMA(5,1,1)(1,0,0)[12]		0,55	-0,03	0,02	0,04	0,06			-0,92	0,01			
SARIMA(5,0,1)(1,0,0)[12]	42,3	0,51	0,11	0,07	0,09	0,14			0,14	0,01			
ARIMAX(5,1,1)													
ARIMAX(7,0,1)	48,8	1,53	-0,56	0,07	0	0,01	-0,03	-0,02	-0,91			-0,01	-0,15
ARIMAX(7,1,1)		0,57	-0,02	0,03	0,04	0,06	0,01	0,04	-0,95			-0,01	-0,15
ARMAX(5,1)	48,8	1,43	-0,5	0,07	0,02	-0,03			-0,81			-0,01	-0,15

	Mean	**AR1**	**AR2**	**AR3**	**AR4**	**AR5**	**MA1**	$\boldsymbol{\omega}$	$\boldsymbol{\alpha_1}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	**Skew**	**Shape**
ARMA(5,1)ARCH(2) 'norm'	43,27	1,42	-0, 49	0,04	0,05	-0,02	-0,80	0,03		0,00	1,00		
ARMA(5,1)ARCH(2) 'std'	43,19	1,35	-0,40	-0,03	0,05	0,02	-0,71	0,06		0,00	1,00		2,71

ARMA(5,1)ARCH(2) 'sstd'	44,96	1,39	-0,42	-0,04	0,06	0,00	-0,76	0,06		0,00	0,99	1,13	2,70
ARMA(5,1)GARCH(1,1) 'norm'	41,97	1,21	-0,28	-0,03	0,07	0,02	-0,64	1,17	0,17	0,77			
ARMA(5,1)GARCH(1,1) 'std'	41,51	1,20	-0,28	-0,04	0,05	0,05	-0,62	1,11	0,17	0,78			5,16
ARMA(5,1)GARCH(1,1) 'sstd'	43,24	1,21	-0,28	-0,04	0,05	0,04	-0,63	1,15	0,17	0,78		1,15	5,09
	alpha0_1	**alpha1_1**	**beta_1**	**nu_1**	**alpha0_2**	**alpha1_2**	**beta_2**	**nu_2**	**P_1_1**	**P_2_1**			
RSGARCH	94	0.9999	0.0000	100	94	0.9999	0.0000	100	0.1256	0.0006			

Post C19	**Mean**	**AR1**	**AR2**	**AR3**	**AR4**	**AR5**	**MA1**	ω	$\propto_1$	β_1	β_2	**Skew**	**Shape**
ARIMA(1,0,0)	19,8	0,37											
ARIMA(5,0,1)	19,8	0,47	-0,27	-0,04	-0,04	0,2			-0,01				
ARIMA(5,1,1)		0,49	-0,26	-0,02	-0,03	0,22			-1				
ARIMA(7,0,1)	19,4	1,34	-0,66	0,19	-0,02	0,2	-0,12	-0,11	-1				
ARIMA(7,1,1)		-0,39	-0,6	-0,55	-0,48	-0,19	-0,09	-0,11	-0,02				
SARIMA(5,1,1)(1,0,0)[12]		0,5	-0,29	-0,01	-0,05	0,22			-1	-0,13			
SARIMA(5,0,1)(1,0,0)[12]	19,7	0,58	-0,37	0,01	-0,06	0,19			-0,11	-0,16			
ARIMAX(5,1,1)		-0,75	-0,6	-0,56	-0,48	-0,21			0,45			0,04	-0,18
ARIMAX(7,0,1)	23,8	0,38	-0,12	-0,06	-0,04	0,12	0,07	-0,13	0,2			0,02	-0,18
ARIMAX(7,1,1)		-0,25	-0,46	-0,38	-0,35	-0,1	-0,01	-0,11	-0,04			0,04	-0,18
ARMAX(5,1)	23,9	0,5	-0,21	-0,02	-0,04	0,15			0,08			0,02	-0,18
	Mean	**AR1**	**AR2**	**AR3**	**AR4**	**AR5**	**MA1**	ω	$\propto_1$	β_1	β_2	**Skew**	**Shape**
ARMA(5,1)ARCH(2)	20,00	0,53	-	-	0,00	0,17	-	0,00		0,98	0,00		
ARMA(5,1)ARCH(2)	22,15	0,50	0,04	-	0,14	0,22	-	1,65		0,82	0,00		2,10
ARMA(5,1)ARCH(2)	20,30	0,31	-	-	0,21	0,16	-	0,00		0,46	0,52	0,59	2,57
ARMA(5,1)GARCH(1,1)norm	20,00	0,53	-	-	0,00	0,17	-	0,00	0,00	0,98			
ARMA(5,1)GARCH(1,1)std	22,15	0,50	0,04	-	0,14	0,22	-	1,65	0,00	0,82			2,10
ARMA(5,1)GARCH(1,1)sstd	19,84	0,68	-	0,03	0,11	0,05	-	0,07	0,04	0,96		0,01	3,06
	alpha0_1	**alpha1_1**	**beta_1**	**nu_1**	**alpha0_2**	**alpha1_2**	**beta_2**	**nu_2**	**P_1_1**	**P_2_1**			
RSGARCH	118	0.9999	0.0000	100	118	0.9999	0.0000	1.000	0.1962	0.0027			

Table 16: Test statistics residuals daily ARMA-GARCH models

Pre-	ARMA(5,1)GARCH(1,		ARMA(5,1)GARCH(		ARMA(5,1)GARCH(		ARMA(5,1)GARCH(0,		ARMA(5,1)GARCH(		ARMA(5,1)GARCH(	
Pearson Goodness of Fit (adjusted)												
	Stat.	P-	Stat.	P-	Stat.	P-	Stat.	P-	Stat.	P-	Stat.	P-
Grou	64,12	0,00	24,76	0,17	11,19	0,92	154,00	0,00	33,58	0,02	19,01	0,46
Grou	85,38	0,00	29,87	0,42	18,94	0,92	174,60	0,00	34,51	0,22	29,41	0,44
Grou	93,10	0,00	37,22	0,55	25,77	0,95	199,60	0,00	57,01	0,03	38,40	0,50
Grou	105,06	0,01	43,02	0,71	37,78	0,88	215,60	0,00	54,07	0,29	37,40	0,89
Ljung- Box on standardized residuals												
Lag[1	1.351	0,2451	0,2207	0,6385	1	0,4135	0,02026	0,8868	0,6823	4,09E+02	0,2516	0,6159755
Lag[5	7.769	0,9843	60.895	10	64	10	609	10.000	139.702	2,83E-	113.578	0,0001127

Lag[9	16.753	0,2584	148.696	0,4924	152	0,449	1.437.221	0,562	228.058	8,19E+00	196.329	0,0635083
d,o,f=												
Ljung- Box on standardized squared residuals												
Lag[1	0,26	0,61	0,36	0,55	0,41	0,52	96	0	82,83	0	75,78	0
Lag[5	0,83	0,90	0,83	0,90	0,92	0,88	166	0	151,92	0	136,71	0
Lag[9	17,48	0,93	16.960	0,94	17,38	0,93	218,9	0	199,84	0	179,58	0
d,o,f=												
Weighted ARCH-LM test												
Lag[3	0,31	0,58	0,25	0,62	0,29	0,59	44,58	0,00	39,10	0,00	34,44	0,00
Lag[5	0,67	0,83	0,58	0,86	0,60	0,85	83,52	0,00	70,16	0,00	63,64	0,00
Lag[7	0,94	0,92	0,91	0,93	0,90	0,93	113,52	0,00	99,35	0,00	89,35	0,00
Post -C19 (Daily)												
Pearson Goodness of Fit (adjusted)												
Grou	15,50	0,69	12,40	0,87	23,60	0,21	15,50	0,69	12,40	0,87	9,20	0,97
Grou	19,87	0,90	19,60	0,90	37,60	0,13	19,87	0,90	19,60	0,90	14,80	0,99
Grou	32,65	0,75	27,15	0,92	57,20	0,03	32,65	0,75	27,15	0,92	20,40	0,99
Grou	56,00	0,23	49,00	0,47	70,00	0,03	56,00	0,23	49,00	0,47	32,00	0,97
Ljung -Box on standardized residuals												
Lag[	0,01114	0,9159	7	7,66E+00	1.739	0,1873	0,01114	0,9159	7.111	7,66E+00	4,29	3,83E+01
Lag[	353	10.000	15	1,11E-	6.550	10	352.649	10.000	14.820	1,11E-	12,27	3,38E-
Lag[	701	0,9997	20	6,44E+01	9.028	0,991	701.172	0,9997	19.609	6,44E+01	16,26	3,13E+02
d,o,f												
Ljung -Box on standardized squared residuals												
Lag[	3.451	0,0632	11,30	0,00	2,48	0,12	3,45	0,06	11,30	0,00	11,65	0,00
Lag[	3.525	0,3194	11,70	0,00	2,57	0,49	3,53	0,32	11,70	0,00	11,92	0,00
Lag[	3.729	0,635	12,01	0,02	2,67	0,81	3,73	0,64	12,01	0,02	12,21	0,02
d,o,f												
Weighted ARCH-LM test												
Lag[	0,00	0,97	0,16	0,69	0,01	0,92	0,00	0,97	0,16	0,69	0,08	0,78
Lag[	0,10	0,99	0,38	0,92	0,04	1,00	0,10	0,99	0,38	0,92	0,28	0,95
Lag[	0,26	0,99	0,50	0,98	0,11	1,00	0,26	0,99	0,50	0,98	0,39	0,99

Table 17: Markov/Regime Switching GARCH model estimates and (state) probabilities

Daily	
Pre-C19 period	**Post-C19 period**

	Coef	Sign	Coef	Sign
α_{0_1}	682.257	**	113	**
α_{1_1}	0,9999	***	0,9999	***
β_1	0	***	0	***
α_{0_2}	686.178	*	114	*
α_{1_2}	0,9999	***	0,9999	***
β_2	0	***	0	***
P_{1_1}	0,9994	*	0,1962	*
P_{2_1}	0,8744	**	0,0027	**

Where 'Coef' stands indicates coefficient and 'Sign' for significance.

State 1 prob	0,9993	State 1 prob	0,0034
State 2 prob	0,0007	State 2 prob	0,9966

	t+1|k=1	t+1|k=2		t+1|k=1	t+1|k=2
t|k=1	0,9994	0,0006	t|k=1	0,1962	0,8038
t|k=2	0,8744	0,1256	t|k=2	0,0027	0,9973

References

Aggarwal, S.K., Saini L.M., Kumar A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems*. 31(1):13–22.

Amjadi, N., and Hemmati, M. (2006). Energy price forecasting. *IEEE Power and Energy Magazine*, March/April, 20-29.

Askari, H. 2008. Oil Price Dynamics (2002-2006). *Energy Economics*. Vol. 30, issue 5, 2134-2153.

Bollersev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3). 307-327.

Boogert, A., & Dupont, D. (2008). When supply meets demand: the case of hourly spot electricity prices. *IEEE Transactions on power Systems*, 23(2), 389-398.

Diebold, F.X. (1986), Modeling the Persistence of Conditional Variances: A comment. *Econometric Reviews*, 5, 51-56.

Diebold F.X. & Mariano, R.S. (1995) Comparing Predictive Accuracy, *Journal of Business & Economic Statistics*, 13:3, 253-263.

Duffie, D., & Gray, S. (1995). Volatility in Energy Prices, *Managing Energy Price Risk*, pp 39-56.

Ethier, R., Mount, T. (1999). Estimating the volatility of spot prices in restructured electricity markets and the implications for option values, *Working paper, Cornell University*.

Garcia-Martos, C., Conejo, A.J. (2013). Price forecasting techniques in power systems. *Wiley encyclopedia of electrical and electronics engineering*, 1-23.

Fong, W.M., & See, K.H. (2002), A Markov switching model of the conditional volatility of crude oil futures prices, *Energy Economics*, 24, issue 1, p. 71-95

Geman, H. & Roncoroni, A. (2006). Understanding the fine structure of electricity prices. *Journal of Business*. Vol. 79 (3), 1225-1261.

Guthrie, G., Videbeck, S. (2007). Electricity spot price dynamics: Beyond financial models, *Energy Policy*, Volume 35, Issue 11, 5614-5621, ISSN 0301-4215.

Haas, M. (2004), A New Approach to Markov-Switching GARCH Models, *Journal of Financial Econometrics*, 2, issue 4, p. 493-530

Haghi, H.V., Tafreshi, S.M.M. (2007). Modeling and forecasting of energy prices using non-stationary Markov models versus stationary hybrid models including a survey of all models. *IEEE Canada EPC* 2007, 429-434.

Hamilton, J., James, D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.

Hansen, P.R., Lunde, A., & Nason, J.M. (2011). The model confidence set. *Econometrica*, Vol. 79, No. 2, pp 453-497.

Huisman, R., Mahieu, R. (2003). Regime jumps in electricity prices. *Energy Economics* 25, 425–434.

Huisman, R., (2008). The influence of temperature on spike probability in Day-Ahead power markets. *Energy Economics*. 30, 2697–2704.

Huisman, R., (2009). An Introduction to Models for the Energy Markets: The Thinking behind

Econometric Techniques and Their Application. *London: RISK Books*.
Huisman, R., Huurman, C., & Mahieu, R. J. (2007). Hourly electricity prices in Day-Ahead markets. *Energy Economics*, 29(2), 240-248.
Geman, H. & Roncoroni, A. (2006), Understanding the Fine Structure of Electricity Prices, *The Journal of Business*, 79, issue 3, p. 1225-1262
Johnsson, T., Pinson,P., Nielsen, H.A., Madsen H., & Nielsen, T.S. (2013). Forecasting electricity spot prices accounting for wind power predictions. IEEE Transactions on Sustainable Energy, 4(1), 210-218.
Kanamura, T., & Ohashi, K. (2008). On transition probabilities of regime switching in electricity prices. *Energy Economics*, 30, 1158-1172.
Karakatsani, N., & Bunn, D. W. (2010). Fundamental and Behavioural Drivers of Electricity Price Volatility, *Studies in Nonlinear Dynamics & Econometrics*, *14*(4).
Knittel, C.R., Roberts, M.R., (2005). An empirical examination of restructured electricity prices. *Energy Economics* 27, 79–817.
Lama, A., Jha, G., Paul, R.K. & Gurung, B. (2015). Modeling and Forecasting of Price Volatility: an Application of GARCH and EGARCH Models. *Agricultural Economics Research Review*, 28(1).
Lamoureux, C., Lastrapes, W. (1990). Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. Journal of Finance, Vol. 45, No. 1, 221-229.
Lubchenco, J., Karl, T.R. (2012). Predicting and managing extreme weather events. *Physics Today* 65(3): 31.
Lucia, J., Schwartz, E.S. (2002). Electricity prices and power derivatives: evidence from the Nordic power exchange. Rev. Derivatives Res. 5, 5–50.
Ma, &., Luh, P., Kasiviswanathan, E. (2004). A neural network-based method for forecasting zonal locational marginal prices. *IEEE Power Engineering Society General Meeting*, conference 6-10 June 2004. ISBN: 0-7803-8465-2.
Mount, T., Ning, Y. & Cai, X. (2006). Predicting prices in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, 28, 62-80.
Olsson, M., Söder, L. (2008). Modeling real-time balancing power market prices using combined SARIMA and Markov processes. *IEEE Transactions on Power Systems*, Vol. 23, no 2, p. 443-450
Raviv, E., Bouwman, K. & van Dijk, D. (2015). Forecasting Day-Ahead Electricity Prices: Utilizing Hourly Prices. *Energy Economics*. 50. 227-239.
Sadorsky, P. (2006), Modeling and forecasting petroleum futures volatility, *Energy Economics*, 28, issue 4, p. 467-488,
Wang, Y.D. & Wu, C.F. (2012). Forecasting Energy Market Volatility Using GARCH Models: Can Multivariate Models Beat Univariate Models? *Energy Economics*, 34, 2167-2181.
Schwartz, E., Smith, J.E. (2000). Short-Term Variations and Long-Term Dynamics in Commodity Prices. *Management Science*, vol. 46(7), 893-911.
We, Y., Wang, Y. & Huang, D. (2010). Forecasting Crude Oil Market Volatility: Further Evidence using GARCH-class models. *Energy Economics*. Vol. 32(6), 1477,1484.

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).

Wolak, F. (1997). Market design and price behavior in restructured electricity markets: an international comparison. *Working Paper. Department of Economics Stanford University*.

Zareipour, H. (2008). Price-based energy management in competitive electricity markets. *VDM Verlag Dr. Müller*.