Erasmus School of Economics
ERASMUS UNIVERSITY ROTTERDAM

Master Thesis Financial Economics

The Predictability of Electricity Spot Prices: An Assessment of the Dutch Market

Candidate
Jan Willem van Os BSc
434162

Supervisor
Dr. R. Quaedvlieg

Second Assessor
Dr. S. van den Hauwe

July 11, 2019

Abstract

The main objective of this study is to examine the predictability of Dutch electricity day-ahead prices by identifying the optimal forecasting model within the AR(F)IMA model family, optionally extended with various dynamics. Based on a recent data set ranging from 2009 up to 2018, evidence suggests that adding day-of-the-week dummies significantly improves forecasts, whereas incorporating month-of-the-year dummies and normalizing price spikes do not. Day-ahead prices are subject to a high-order autocovariance structure. As a result, the HAR model with weekly and monthly dependency, H(7,30)AR(1), and day-of-the-week dummies is found optimal. Next to that, this study is among the first to analyze the relationship between the day-ahead market and the imbalance market. In particular, a negative relationship between day-ahead price predictability and imbalance price volatility is documented. These results have various implications for traders on both power markets, and contribute to the academic foundation on Dutch prices.

Keywords: Dutch day-ahead market, Dutch imbalance market, AR(F)IMA price forecasting, time fixed effects, spikes.

JEL-codes: C22, C51, C52, C53.

Contents

1 Introduction

One of the most prominent topics in modern economies is sustainability. In particular, carbon dioxide emissions are considered a major threat to the climate. Induced by governmental policies, such as the European strategy targets (European Union, 2009), and societal factors, many countries aim towards a reduction of the use of fossil fuels to generate energy. In this pursuit, electricity is a favorable alternative, as it can be generated using renewable sources, such as wind and sun. However, electricity is subject to one major problem: it cannot be stored efficiently—only on a relatively small scale. The economic consequence of this property is that in an efficient and reliable electricity market, supply and demand should match continuously. But both supply and demand suffer from shocks and imperfect predictability, leading to many frictions and challenges. As the share of renewable sources increases, aggregated supply becomes even more unpredictable and volatile. Consequently, electricity markets are structured in a complex manner to provide a reliable grid system.

Most electricity is traded between wholesale parties on the day-ahead market, that is typically classified as the spot market. For parties that trade on this market, price predictions are crucial in their bidding strategies. Most importantly for profit maximization or reducing risk on consumption or production (Tan, Zhang, Wang, & Xu, 2010). But mainly due to the non-storability property, prices on this market are subject to complex and unique patterns, such as systematic price differences between days of the week and months of the year, strong price spikes and time-varying volatility.

Although large players on this market employ their own customized and optimized models, forecasting models are not that accessible for new entrants or start-ups. A fundamental academic basis describing electricity price dynamics is therefore crucial. Many studies have focused on price forecasting within a wide range of power markets. However, there is a substantial academic gap as to the Dutch day-ahead market. The main objective of this study is therefore to address the day-ahead price predictability in the Dutch market. Using a recent price set from 2009 up to 2018, a range of AR(F)IMA time series models are extended with expected price dynamics aimed towards optimizing the forecasts. The goal is to identify the optimal predictive model, and identifying relevant price patterns in the process.

To provide a reliable power transmission system, the imbalance market acts as a balancing mechanism. In this market, real-time shortages or surpluses of electricity are balanced by means of a semi-economic mechanism (TenneT Holding B.V., 2016). Naturally, this market is subject to

strong volatility driven by heavy supply and demand shocks. As such, it provides opportunities for emerging business models, such as "smart" appliances detecting whether electricity is cheap or expensive at a certain moment in time. But as of yet, studies analyzing this market are scarce. Logically, the imbalance market is expected to be related to the predictability component on the day-ahead market. More specifically, when day-ahead prices are relatively unpredictable, the imbalance market is expected to be subject to more or stronger shocks. This study is among the first to assess this relationship in an empirical manner, using a variety of measures.

The research question is formulated according to the two described components.

Research Question *To what extent are Dutch electricity day-ahead prices predictable and how does predictability relate to intra-day imbalance trading?*

Using an out-of-sample period of two years (2017 and 2018), in combination with an in-sample rolling window of eight years, the main conclusion is that an H(7,30)AR(1) model with day-of-the-week dummies is optimal in a predictive context, based on both performance and parsimony. The mean absolute percentage error is equal to 7.59%, relatively in line with other studies. Based on different specifications, the day-of-the-week effect clearly enhances forecasts, whereas incorporating month-of-the-year dummies does not. Pre-filtering the prices for spikes does not yield significant improvements either. The comprehensive analysis of various forecasting models is very helpful for emerging businesses that seek optimal predictions on the day-ahead market. As for the imbalance market, a negative relationship was found between intra-day price volatility and the predictability of day-ahead prices. For investors, this implies that positive profits can be achieved by trading on the imbalance market when day-ahead prices are relatively unpredictable. However, insufficient evidence was found for a negative relationship to exist between day-ahead price predictability and the daily imbalance market volume. Trading parties presumably adjust their capacity rapidly after supply and demand shocks, minimizing the impact on the daily aggregated volume.

The remainder of this study is structured as follows. Chapter 2 develops a theoretical basis for the research question and formulates the theoretical hypotheses. Subsequently, in Chapter 3, the analyzed data set is presented and described in detail. Chapter 4 elaborates on the methods that are applied on the data set in order to arrive at relevant results. The prediction results are presented and discussed in Chapter 5 and Chapter 6 documents the analysis on the imbalance market. Finally, the research question is evaluated in Chapter 7, along with a discussion and recommendations for future studies.

2 Theory

2.1 Fundamentals

Electricity represents a unique commodity class due to its special characteristics. Most importantly, the current state in technology does not provide a way of storing electricity efficiently and economically. As a result, trading electricity comes with many complexities relative to other commodities. Because of this non-storability property, demand and supply must match at all times. Surpluses cannot be deferred for consumption later. At the same time, shortages can induce dramatic economic and societal consequences. As a consequence, sophisticated market structures are required to provide a reliable power system. This section develops a foundation for the study by elaborating on the Dutch market structure and its key dynamics.

In the Netherlands, TenneT is the transmission system operator (TSO) for electricity and the institution that is responsible for a reliable grid system. Electricity is generally traded between wholesale parties, distributed over four different markets, each representing different purposes and dynamics (TenneT Holding B.V., 2018). Firstly, for long horizons, parties can buy and sell power via standardized futures contracts. Secondly, the day-ahead market provides a trading base for the delivery of power for the next day. The intra-day market further allows for high-frequency hourly trading. Finally, the imbalance market provides a way of dealing with mismatches that cannot be assimilated in the other markets.

The main objective of this study is to assess the predictability of day-ahead prices, that are typically used as a proxy for spot prices. Therefore, the day-ahead market is introduced in more detail. The secondary part of the research question concerns the relationship between the day-ahead price predictability and the imbalance market. Consequently, the fundamentals of the imbalance market are elaborated on as well.

2.1.1 Day-ahead market

Along with rapidly improving technology, many countries have been liberalizing and deregulating their electricity sectors, generally aimed at introducing incentives to innovate and to improve efficiency (Weron, 2006, p. 1). As a result, trading electricity has become easier and more accessible. Many commodities trade in a spot market in which parties trade instantaneously. However, as electricity is a sensitive commodity that requires a high level of reliability, the trading mechanism is characterized by a one-day delay so that the TSO can effectively manage the transmission system (Weron, 2006, p. 7). Nevertheless, day-ahead prices are typically referred

to as electricity spot prices. European day-ahead markets have been becoming more integrated in the recent decades, a development that is referred to as market coupling (Zachmann, 2008; Huisman & Kiliç, 2013). As of 2015, trading on the Dutch day-ahead market is centralized at the European Power Exchange (EPEX) on the EPEX spot market (TenneT Holding B.V., 2018). Therefore, day-ahead electricity prices are referred to as EPEX spot prices. Trading on the day-ahead market is not restricted to trading hours only, the market is active every hour of every day. Up to 12 AM, market participants can file their orders for delivery on the next day for all hours simultaneously into the EPEX market system. Subsequently, demand and supply are matched and an equilibrium price is constructed automatically (EPEX SPOT SE, 2018). Based on the established price, market participants execute (a portion of) their orders.

Electricity day-ahead prices have been investigated extensively. For several European countries, Huisman, Huurman, and Mahieu (2007) demonstrate that hourly day-ahead prices do not form an aggregated time series. Instead, they form a panel of individual time series indicated by a specific mean and volatility distribution, implying that prices differ between hours and that hourly prices are related inter-daily. Even though this suggests that hourly prices contain valuable information, most studies focus on base load prices, computed as daily 24-hour averages (Huisman et al., 2007). This is however justified, as base load prices are broadly employed as proxies for daily spot prices and referencing instruments for derivatives (Raviv, Bouwman, & van Dijk, 2015).

The structure of day-ahead markets depends on particular fundamentals. Especially due to the non-storability property, dynamics of supply and demand play a central role in modelling electricity markets. The supply curve is typically driven by the merit order effect, where power production processes are deployed based on their marginal production costs to match demand (Mulder & Schoonbeek, 2013; Clò, Cataldi, & Zoppoli, 2015). As a result, the supply curve exhibits a convex step-by-step structure. Electricity demand is, at least on the short term, close to being perfectly inelastic (Borenstein, 2002). Due to these dynamics, established electricity prices are determined by the marginal cost of the most expensive energy source needed to fulfill demand.

As a consequence of these particular market dynamics, day-ahead prices are subject to specific patterns. Most series of electricity spot prices are described by a stationary process, implying that statistical methods can be employed effectively (Knittel & Roberts, 2005). Aligned with stationarity, electricity spot price processes are typically mean-reverting and therefore sometimes referred to as anti-persistent (Weron, 2008). However, electricity prices are also found to contain high-order autocovariance structures and thereby exhibit long-memory behavior (Haldrup

& Ørregaard Nielsen, 2006). In other words, due to a relatively long-term autocorrelation, prices are clustered in wide intervals. Another key property is the seasonal dependence along different dimensions (Knittel & Roberts, 2005). On the demand side, economic activity as well as weather conditions are important drivers. Supply may also be influenced by weather properties, conditional on the production mix that is employed. These seasonality effects manifest in differences in mean and volatility distributions between days of the week and months of the year (Weron, 2008). Finally, electricity spot prices typically suffer from infrequent extremes, generally referred to as spikes. In contrast to equity markets in which jumps typically persist, spikes in electricity markets quickly die out and revert back to the local mean (Huisman & Mahieu, 2003). Price spikes occur as a result of demand and supply shocks in the short run, for example as a result of extreme weather conditions (Huisman, 2008), and cause electricity prices to exhibit an extreme volatility structure, relative to prices of various financial products and other energy-related commodities (Weron, 2006, p. 26).

2.1.2 Imbalance market

Due to the non-storability property of electricity and the fact that supply and demand cannot be predicted perfectly, the futures, day-ahead and intra-day markets are not sufficient to provide a reliable grid system. In that regard, the imbalance market (or balancing market) provides a last resort for mismatches. Most importantly, it provides a safety net for incurred shortages of electricity, that can be compensated by other Balance Responsible Parties (BRPs) or reserves held by the operator. On the other hand, overproduction can be sold through this market, benefiting BRPs. The existence of this market essentially transforms the balancing problem to a structured economic process. The market is managed by the TSO, TenneT, who acts as an administrative counterparty in all transactions (TenneT Holding B.V., 2016). The importance of the imbalance market is highlighted in different studies. As the share of renewable energy sources increases, the role of the imbalance market becomes more important as an instrument to provide reserves (Purvins, Zubaryeva, Llorente, Tzimas, & Mercier, 2011). As a result, Farahmand and Doorman (2012) argue that integrating imbalance markets in Northern Europe would improve the aggregated efficiency.

In contrast to the day-ahead market where orders are hourly-based, the imbalance market is structured along Program Time Units (PTUs) that represent windows of 15 minutes (TenneT Holding B.V., 2016). For every PTU, the market encompasses supply and demand, combined creating an imbalance that represents either a surplus or a shortage. Although a specific pricing scheme applies, for example penalties that create an incentive to minimize the system imbalance (TenneT Holding B.V., 2016), dominating prices are simply driven by the relative equality of

supply and demand. More specifically, the occurring imbalance is negatively correlated with the established price. This implies that when there is more supply than demand, electricity is cheap and vice versa. As a result, trading on the imbalance market can be employed as a profitable strategy by providing short-term storage. The most prominent and topical application is the introduction of "smart" appliances—for example electrical vehicle charging stations known as vehicle-to-grid (Gough, Dickerson, Rowley, & Walsh, 2017)—that detect whether electricity is cheap or expensive at specific moments.

2.2 Literature review

In building predictive models, it is key to incorporate relevant dynamics and correct assumptions. Therefore, and based on the previous sections describing the fundamentals of electricity markets, this section reviews the current state of academic literature regarding forecasting frameworks. The goal is to develop a theoretical foundation of the predictive approaches that have been applied through time to base both the hypotheses and methodology on.

As a result of the deregulation and integration of many electricity markets, forecasting exercises have been employed broadly since approximately 20 years ago. Nogales, Contreras, Conejo, and Espinola (2002) and Contreras, Espinola, Nogales, and Conejo (2003) analyze the Spanish and Californian markets and were among the first to successfully introduce time series models in day-ahead forecasting based on the identified dynamics in these markets. Although these studies analyze only a limited horizon of electricity prices of only a few weeks on an hourly basis, they find relatively small one-digit average daily errors, suggesting that utilizing time series models in predictive frameworks is attractive. Cuaresma, Hlouskova, Kossmeier, and Obersteiner (2004) further extend this approach in a day-ahead context and analyze hourly electricity prices from the Leipzig Power Exchange. Next to a broader selection of models, they explicitly incorporate day-of-the-week dummies, month-of-the-year dummies and an additional factor describing price spikes. Their results clearly suggest presence and predictive power of seasonality components, and allowing for spikes sometimes induces marginal improvements. In terms of absolute performance, they observe mean average errors between approximately 3.2 and 7.1 and root mean squared errors between 4.9 and 10.0. Kristiansen (2012) similarly applies an autoregressive specification on hourly prices, incorporating the day-of-the-week effect by including dummies. As an extension, both wind power and electricity demand are included as exogenous explanatory variables, which turn out to have a significant effect on the established price. Hourly mean absolute percentage errors of approximately 5% are obtained. More recently, Raviv et al. (2015) employ predictive models on prices from the Nordic power market, incorporating both the day-of-the-week effect and the month-of-the-year effect. In addition, they propose heterogeneous

autoregressive (HAR) models to allow for the long-memory property. In forecasting base load prices, they observe mean average errors ranging from 9.5 to 11.9, as well as root mean squared errors between 19.7 and 23.4. Overall, it is evident that studies yield very different results, mainly as a result of differences in methodology and sample selection.

Through time, many extensions have been applied to basic time series models to correct for electricity-specific dynamics. Garcia, Contreras, van Akkeren, and Garcia (2005) introduce GARCH modelling to deal with heteroskedasticity in the volatility distribution in ARMA processes. Subsequent studies followed this approach (Bowden & Payne, 2008; Liu & Shi, 2013). Other studies employ wavelet transformations to obtain more convenient price subsets that are more suitable for individual models (Conejo, Plazas, Espinola, & Molina, 2005; Tan et al., 2010), use other techniques such as artificial neural networks (Amjady, 2006) or include exogenous parameters to provide additional predictive power (Kristiansen, 2012).

The most apparent difficulty in econometric electricity models is the existence of price spikes. Many studies focus on explicitly modelling these. Clewlow and Strickland (2000) introduce the Merton (1976) jump diffusion model as an instrument to explain the behaviour of extreme prices. Huisman and Mahieu (2003) complement the stochastic jump diffusion model class by introducing switching regimes, and document substantial improvements as mean reversion is more effectively incorporated. From a forward-looking perspective, Mount, Ning, and Cai (2006) develop a framework to predict the occurrence of spikes based on one-day ahead forecast loads. However, their model is heavily sensitive to the availability and reliability of external information. As a consequence, the existence of spikes still represents a considerable problem in price forecasting models, most importantly as their occurrence is hard to predict.

In a predictive context, most studies employ a two-step pre-processing methodology, where spikes are first identified and subsequently normalized before processed in statistical frameworks (Janczura, Trück, Weron, & Wolff, 2013). Boogert and Dupont (2008) classify any price above a fixed threshold as a spike. Other studies construct recursive variable thresholds for either prices or returns based on confidence intervals around the local or global mean and, typically, three times the standard deviation (Weron, 2008; Keles, Genoese, Möst, & Fichtner, 2012). Normalizing the spikes is however much more complex. The fundamental problem is the absence of a definition for a normal price (Weron, 2008). Different treatments include limiting to the relevant boundary or replacing with surrounding prices (Weron, 2006, p. 167). Shahidehpour, Yamin, and Li (2002, p. 83) further prescribe a dampening procedure based on a logarithmic transformation, having a milder effect on spikes and thereby leaving the price structure more intact. As of yet, there is no consensus as to what spike pre-processing approach works best.

Complementary to differences in dynamics between distinct electricity markets, data frequency (hourly or daily) as well as forecasting horizons can have a substantial impact on the effectiveness of pre-filter algorithms in predictive frameworks.

2.3 Hypotheses

Currently, modern countries are shifting towards more sustainability, prompted by government policy and driven by societal factors. A major consequence is that electricity is becoming more important as a source for energy, incrementally replacing fossil fuels. In conjunction with evolving electricity markets, it is economically crucial to develop a broad and strong empirical foundation regarding dynamics that currently manifest in electricity prices. Furthermore, being able to forecast prices is especially relevant for emerging businesses trying to develop business models around the energy transition.

Literature around electricity price forecasting is relatively broad. Due to different pricing characteristics, time series models clearly represent reasonable instruments for predictive purposes. However, there is an evident empirical gap regarding the Dutch market. Although some studies have focused on identifying patterns in Dutch electricity prices, the predictability is not yet examined. The present study attempts to fill that gap, by applying AR(F)IMA time series models on Dutch electricity spot (day-ahead) prices. In that attempt, it is crucial to assess which pricing dynamics are present in the utilized data set and, more importantly, which factors improve forecasting performance. Various dynamics are observed in academic literature. Therefore, several theoretical hypotheses are established representing these dynamics and their expected effect on predictive models. Based on these hypotheses, one or more optimal models will be selected.

Firstly, electricity prices typically differ between days of the week. Incorporating this in predictive models is consequently expected to improve forecasts. This expectation is reflected in the first hypothesis.

Hypothesis 1 *Adding a day-of-the-week effect improves the forecasting performance for Dutch day-ahead electricity prices.*

Additionally, due to changing weather conditions, day-ahead prices are subject to changes between different months of the year. During summer and winter, prices are typically highest. Accounting for this is expected to have a positive effect on forecasting approaches, and is materialized in the second hypothesis.

Hypothesis 2 *Incorporating a month-of-the-year dependency improves statistical forecasting power for Dutch electricity spot prices.*

Various studies note that electricity spot prices suffer from high-order autocovariance structures. Incorporating this dynamic requires specific models, and is expected to be present in Dutch prices as well. The third hypothesis is defined as follows.

Hypothesis 3 *The forecasting accuracy of Dutch electricity spot prices improves by accounting for a long-memory feature.*

Price spikes are also documented to have a substantial impact on predictive frameworks. Combined with mean-reversion, filters can be developed to deal with these spikes ex ante. Correcting for these spikes is therefore expected to positively affect predictive power.

Hypothesis 4 *Dutch day-ahead electricity price predictions improve when spikes are filtered.*

Related to day-ahead predictability, the imbalance market represents striking dynamics. Fluctuating demand and supply of electricity on the short term (within the day) causes day-ahead quantity predictions to be imperfect. The imbalance market is then needed to compensate for surpluses or shortages on a real-time basis. Since the pricing mechanism on the day-ahead market is based on the intersection between supply and demand, price predictability can be used as a proxy for the predictability of supply and demand. It follows that if day-ahead prices are relatively unpredictable, the imbalance market will be subject to supply and demand shocks. Supply and demand shocks on this market materialize into a large price volatility. This translates into the following hypothesis:

Hypothesis 5 *The predictability of Dutch day-ahead electricity prices relates negatively to price volatility on the imbalance market.*

Lower levels of day-ahead predictability imply that the imbalance market is required to compensate a larger volume. In effect, this would yield greater system imbalances, correlated with greater price deviations, as argued by the previous hypothesis. Additionally, this would imply that total daily volumes would be larger as well. Using this metric is however less reliable, since PTU-specific observations are simply aggregated and individually faded. However, with great price uncertainty, it is expected that the imbalance market exhibits more activity. The final hypothesis is therefore formulated as:

Hypothesis 6 *Daily volumes on the Dutch imbalance market negatively relate to price predictability on the day-ahead market.*

3 Data

3.1 Selection and construction

In every study, selecting appropriate data is crucial. The main objective of this study is to employ statistical models to assess the predictability of electricity spot prices in the Dutch day-ahead market. Initially, 24-hour Dutch EPEX spot price data spanning from January 2000 up until December 2018 is obtained from Bloomberg, resulting in 6,940 days of hourly quotes (166,560 quotes in total). In contrast to securities on stock markets, electricity is traded every hour of every day—not only during trading periods. However, up until July 2000, prices during weekends are not or only partially documented. Consequently, and since time series play the central role in this study, year 2000 is excluded as a whole from the data set.

From 2001 to 2018, a small fraction of observations is missing (approximately 10 days in total). Additionally, some market dynamics—including a summer-winter time transition resulting in one-hour differences between days—led to blanks. These missing observations are simply replaced by the hourly quote of the day before. The adjusted full sample of day-ahead prices consists of 6,574 days ranging from January 1, 2001 up to December 31, 2018 and a total of $6,574 \cdot 24 = 157,776$ hourly prices.

Similarly to related literature studying electricity spot prices, a daily base load price P_t is constructed as the average of the 24 daily quotes (Huisman & Kiliç, 2013; Raviv et al., 2015). That is

$$P_t = \frac{1}{24} \sum_{h=1}^{24} P_{t,h}, \tag{1}$$

where $P_{t,h}$ denotes the quoted price at day t for hour $h = 1, 2, \cdots, 24$.

To test for predictability, it is key to define both an in-sample data range and an out-of-sample range, for estimating models on and evaluating the models' accuracy with, respectively. On the one hand, practitioners should utilize as many observations in the in-sample range to maximize the models' accuracy by using a high number of data points. However, if the time series exhibits great time variance, including obsolete observations reduces the applicability on more recent observations. Raviv et al. (2015) compute out-of-sample forecasts for approximately two years using a rolling window of five years, arguing that a window of five years is sufficient for the construction of valid coefficients while allowing for time variety. Cuaresma et al. (2004) conduct forecasting models using a window of only approximately one year. However, these

studies employ relatively simple models, as compared to the current study. Therefore, the aim is to include as many reliable observations as possible. A background analysis, that is elaborated on in more detail in Section 3.2, has shown that prices before 2009 exhibit very different dynamics than after. To obtain consequent and realistic results, all observations before 2009 are dropped. The last two years of the adjusted sample, 2017 and 2018, are used as an out-of-sample period to provide a representative large number of forecasts, and the eight preceding years as the in-sample subset.

Additionally to the day-ahead market, the imbalance market is analyzed. All intra-day data is obtained from TenneT, the TSO for the Dutch market. Aligned with the day-ahead range, the initial sample consists of intra-day data ranging from January 1, 2009 to December 31, 2018. Each day reports $24 \cdot 4 = 96$ entries (by a 15-minute window) comprising different variables related to market activity. The sample contains 3,652 days and hence $3,652 \cdot 96 = 350,592$ PTU observations. First, buy and sell prices are reported. For convenience, one general imbalance price $P^{\asymp}_{t,u}$, for PTU u at day t is calculated by taking their average. To measure the intra-day volatility of electricity prices, $\sigma^{\asymp}_t$ is calculated as the standard deviation of the imbalance prices. That is

$$\sigma^{\asymp}_t = \sqrt{\frac{1}{96}\sum_{u=1}^{96}(P^{\asymp}_{t,u} - \overline{P}^{\asymp}_t)^2}, \tag{2}$$

where $\overline{P}^{\asymp}_t = 96^{-1}\sum_{u=1}^{96} P^{\asymp}_{t,u}$ and the superscript $\asymp$ indicates a variable related to the imbalance market.

Next to prices, quantities are reported. Let $S^{\asymp}_{t,u}$ and $D^{\asymp}_{t,u}$ denote the supply and demand of electricity on the imbalance market, respectively, for day t and PTU u. Then, at that data point, the imbalance market is constituted by the total volume $V^{\asymp}_{t,u} = S^{\asymp}_{t,u} + D^{\asymp}_{t,u}$. The daily size of the imbalance market is consequently measured by the aggregated volume

$$V^{\asymp}_t = \sum_{u=1}^{96} V^{\asymp}_{t,u}, \tag{3}$$

expressing the extent to which the imbalance market is needed to deal with deficits and surpluses between the production and consumption of electricity. Table 1 contains an overview of the final samples that are utilized in the analysis.

3.2 Day-ahead properties

As discussed in the previous chapter, day-ahead markets typically exhibit specific characteristics such as stationarity, mean reversion, seasonality and autocorrelation. A preliminary analysis of

Market	Sample	Start Date	End Date	Frequency	Observations
Day-ahead	In-sample[1]	01/01/2009	12/31/2016	Daily	2,922
	Out-of-sample	01/01/2017	12/31/2018	Daily	730
Imbalance	In-sample	01/01/2009	12/31/2018	Daily	3,652

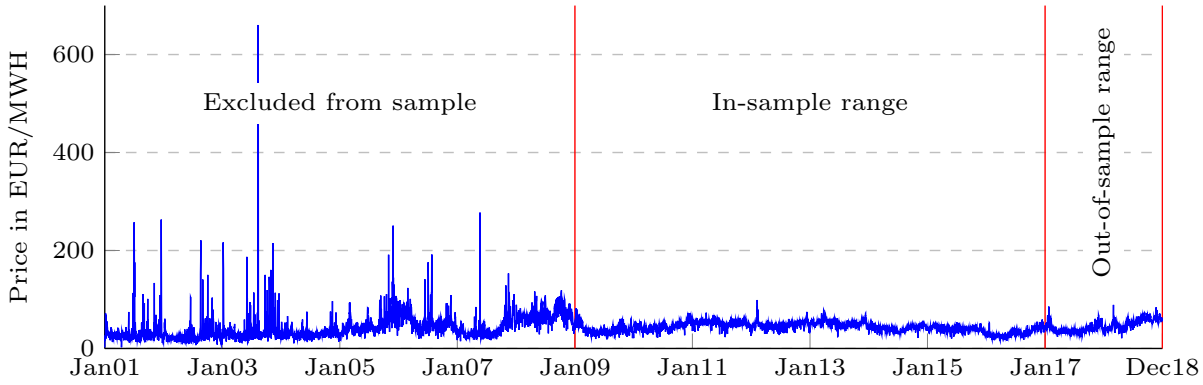
[1] The in-sample set is employed as a rolling window used in forecasting the out-of-sample set.

Table 1: Description of the different data sets employed in the analyses

the selected data will therefore be informative and shed light on the fundamentals of the Dutch electricity market specifically. Further, a careful assessment of the autocovariance structure is needed to match time series models with.

Figure 1 illustrates the average daily electricity price on the EPEX spot market for the full data set, where both the in-sample and out-of-sample range are featured. It is evident that there is a break in volatility and mean distribution around the beginning of 2009. A quantitative analysis supports this claim, and clearly documents different dependencies on, for example, time fixed effects. The more stable structure of electricity prices from 2009 onward is consistent with improved market coupling in the European Union (Zachmann, 2008; de Menezes, Houllier, & Tamvakis, 2016). Further, the price process exhibits a mean reverting effect, consistent with findings of related literature.

Figure 1: The development of Dutch day-ahead prices for the full initial sample



3.2.1 Stationarity

Before working with time series data in forecasting frameworks, it is crucial that the data exhibits a stationary structure. Running estimations on non-stationary data can lead to spurious regressions and therefore invalid relationships (Brooks, 2014, p. 354). As a consequence, forecasts based on a non-stationary series presumably turn out poor. A process that is weakly stationary, which is the standard requirement for time series modelling, generally has a constant mean, variance and autocovariance structure, typically resulting in a plot that hovers around a static

mean.

Figure 2: The development of Dutch day-ahead prices for the select sample

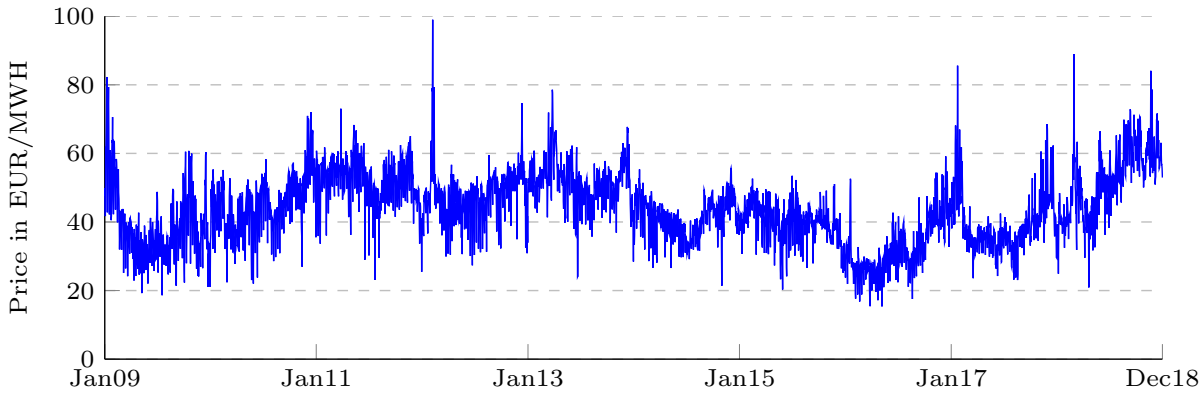


Figure 2 illustrates the day-ahead price development throughout the selected sample starting in 2009. The graph exhibits no clear trend through time. However, it is evident that some periods contain more extremes and higher variance than other periods. Periods with extreme values die out, but slowly. It is therefore key to assess the stationarity using quantitative tests.

Non-stationarity can manifest in different forms. A predominantly encountered problem is the case of unit root, where a time series is stated to be integrated of order 1—denoted as $\sim I(1)$, which implies that the series should be differenced once to become $\sim I(0)$, stationary (Brooks, 2014, p. 362). The presence of a unit root yields a situation in which a time series cannot be modelled, since the expected values are simply a sum of the inherent shocks. The Dickey-Fuller test essentially estimates whether the coefficient is equal to one in a simple AR(1) setting (Dickey & Fuller, 1979). Table 2 presents the results of a Dickey-Fuller test without trend, with a constant. The 1%-significant statistics for both the full sample and in-sample range imply that the null hypothesis of the series containing a unit root is rejected, which suggests that the electricity prices are not integrated with factor 1. The Phillips and Perron (1988) test is included as well, which adds additional power to the Dickey-Fuller test by correcting for potential serial correlation in the errors by incorporating Newey and West (1987) errors (similar to the Augmented Dickey-Fuller test, which allows for one or more lags).

To complement the results of the unit root tests, the KPSS test is employed. Conversely to the unit root tests, the KPSS approach has stationarity in its null hypothesis against the series containing a unit root under the alternative hypothesis (Kwiatkowski et al., 1992). Table 2 contains the results of the test, from which it clearly follows that the null hypothesis of the process following an $I(0)$ process, is rejected.

Although these results suggest the absence of a unit root, stationarity is rejected as well. However, the KPSS test typically fails to identify fractionally integrated stationarity (with

Test	H_0	Statistic		Critical Values		
		Full sample	In-sample	1%	5%	10%
Dickey-Fuller	$P_t \sim I(1)$	-18.785***	-17.472***	-3.430	-2.860	-2.570
Philips-Perron		-17.775***	-16.791***	-3.480	-2.860	-2.570
Kwiatkowski-Philips-Schmidt-Shin	$P_t \sim I(0)$	16.1***	22.6***	0.216	0.146	0.119

Notes: This table represents an assessment of the stationarity of both the full sample and in-sample sets. In particular, results for the Dickey and Fuller (1979), Phillips and Perron (1988) and Kwiatkowski, Phillips, Schmidt, and Shin (1992) tests are documented. All tests are conducted on non-differenced prices (lag 0) and *** denotes 1% significance.

Table 2: Results of stationarity tests for electricity prices

$0 < d < 1$) in its null hypothesis (Lee & Schmidt, 1996). Fractionally integrated series follow an $I(d)$ process where d is not an integer, but another real number. This phenomenon results in observations being clustered with high autocorrelation, such that the dependence is relatively long-term, but not sufficiently to form a unit root. Such series can be modelled in an ARFIMA(p,d,q) specification, as introduced by Hosking (1981). Lee and Schmidt (1996) further emphasize that a series following an $I(d)$ process with $-0.5 < d < 0.5$ is revertible and stationary.

From an ARFIMA(0,d,0) analysis d is estimated as $\hat{d} \approx 0.496 < 0.5$ over the entire sample and $\hat{d} \approx 0.491$ for the in-sample range specifically. Although the hard boundary of 0.5 is not reached, coefficient estimates typically never approach extreme values, which is the precise reason why a Dickey and Fuller (1979) test is required instead of a regular regression to test whether the AR(1) coefficient is equal to one. A widely employed alternative is the transformation into returns, that typically yields a stationary series. In order to assess whether regular prices are reasonably stationary, out-of-sample forecasts are computed for AR(p) models with $p = 1, \cdots, 10$ from the perspective of both prices and returns, of which Table 12 in Appendix A presents the performance criteria. It is evident that modelling returns instead of prices does not improve predictive power. The set of electricity prices is thus assumed a fractionally stationary series—explicitly differencing the series is therefore unnecessary. Further, $\hat{d} < 1$ proves mean reversion (Baillie, 1996). Another relevant implication from this result is that fractionally integrated specifications are likely to outperform similar ARMA(p,q) models, since they fit the data better.

3.2.2 Descriptive statistics

A summary of quantitative statistical measures is provided in Table 3. The day-ahead prices are classified in two dimensions: per day of the week and per month of the year. It is evident that electricity prices differ in days of the week and that electricity is cheapest during the weekends and the most expensive around Wednesdays. The volatility also increases with the mean. Prices

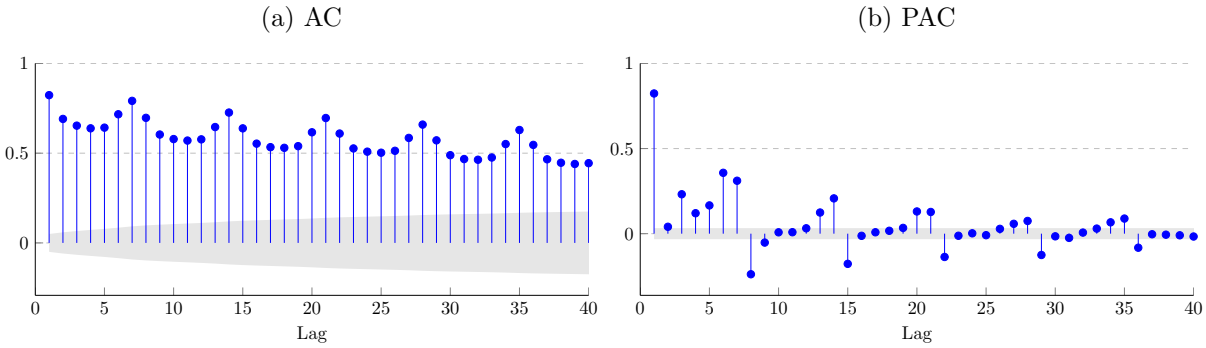
differ between months of the year as well. Electricity is relatively expensive during winter and also during fall. Between April and August, electricity prices are, on average, steady and low, hovering around €41 per megawatt hour. Summers do not seem to have an upward effect on prices, in contrast to the findings of, for example, Knittel and Roberts (2005). The skewness and kurtosis metrics indicate the extent to which the data exhibits a normal distribution. For most subsets, the skewness is far from zero, indicating that prices are not symmetric. Some subsets suffer from heavy tails and others light tails, corresponding to a kurtosis measure larger and smaller than three, respectively. Overall, base load electricity prices cannot be assumed following a normal distribution on a consistent basis.

Subset	Observations	Mean	Stdev	Minimum	Maximum	Skewness	Kurtosis
Panel A: Per day of the week							
Monday	522	45.532	9.973	15.432	79.267	0.191	3.136
Tuesday	521	46.745	9.857	24.710	85.585	0.283	3.035
Wednesday	521	46.553	10.248	23.115	98.982	0.397	3.772
Thursday	522	46.467	10.041	20.392	84.089	0.308	3.178
Friday	522	45.962	9.830	21.042	88.975	0.311	3.570
Saturday	522	41.086	9.013	19.158	67.423	0.118	2.737
Sunday	522	36.934	9.263	15.372	64.771	0.197	2.619
Panel B: Per month							
January	310	45.596	10.443	21.049	85.585	0.462	3.853
February	282	45.690	10.978	16.808	98.982	0.213	5.282
March	310	42.954	11.337	15.432	88.975	0.537	3.683
April	300	41.435	9.935	17.178	66.776	0.140	2.698
May	310	41.216	10.702	15.372	68.299	0.234	2.529
June	300	41.675	9.478	22.717	63.009	0.196	2.175
July	310	41.772	8.736	18.628	65.262	0.029	2.118
August	310	41.525	9.813	17.541	69.826	0.301	2.830
September	300	45.528	9.733	21.379	72.876	0.318	3.025
October	310	47.125	8.376	22.343	70.733	0.102	2.885
November	300	48.130	10.006	21.373	84.089	0.210	3.210
December	310	47.668	10.546	21.042	74.687	0.075	2.528
Aggregated	3652	44.181	10.348	15.372	98.982	0.239	3.225

Table 3: Description of the different data sets employed in the analyses

In predicting day-ahead electricity prices, it is key to assess the autocorrelation between the observations for selecting appropriate forecasting models. As depicted in Figure 3, electricity prices are strongly driven by both autocorrelation and partial autocorrelation. Autocorrelation is present for an unknown number of lags, at least up until lag 40. Partial autocorrelation, which captures the direct correlation with a particular lag k excluding the autocorrelation up until lag $k-1$, is at least present up until the ninth lag. From lag ten onward, it gradually decays with

Figure 3: (Partial) autocorrelation for APX electricity day-ahead prices

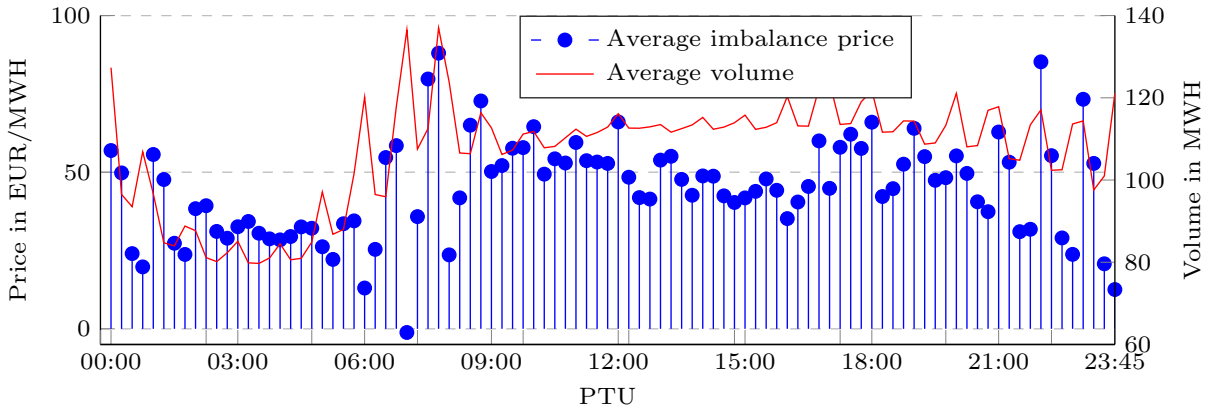


occasional spikes occurring, which confirms the long-memory observation as mentioned earlier. The correlation spikes are likely driven by the day-of-the-week effect.

3.3 Imbalance properties

The imbalance market typically exhibits even more fluctuations than the day-ahead market. In fact, prices can become negative due to the great dispersion between supply and demand. Approximately 6% of PTU-based prices are reported negative from January 2009 up to December 2018. Figure 4 illustrates average prices and volumes for each PTU.

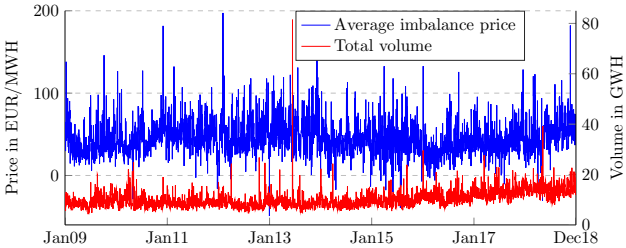
Figure 4: PTU averages of imbalance prices and volumes



During the night, the imbalance market entails less activity and lower price volatility. Prices tend to spike when upward jumps in volume are observed. It further appears that volume shocks are greatest in the evening and between 06:00 and 08:00 in the morning, typically referred to as peak hours. Wholesale parties apparently suffer from supply/demand jumps during these periods. The volume seems relatively stable between 08:00 and 16:00, but hovers around a relatively high level. One explanation for this is that the market for electricity as a whole is much larger during this period, making it easier to provide a cushion for imbalances, while at the same time having more capabilities of adjusting short-term production.

Figure 5: Development of the imbalance market over time

(a) Average prices and volumes



(b) Daily price standard deviation with trend

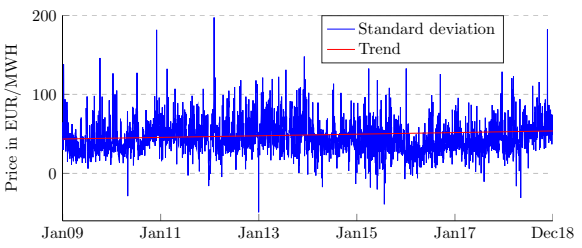


Figure 5a presents the development of the imbalance market over time. Average daily prices and total daily volumes are included. Although one would expect that due to technological advantages electricity supply and demand would be matched more efficiently, the imbalance market provides no evidence for that. Instead, volumes are on average subject to a positive trend, confirmed by a significant coefficient $\hat{\beta} \approx 1.43$ in a regression setting $V_t^{\asymp} = \alpha + \beta t + \varepsilon_t$, which implies that the volume of the imbalance market has increased by approximately 1.43 megawatt hour per day between 2009 and 2018. The predominant explanation is the increased share of renewable energy sources in electricity production, which induces supply volatility. This explanation is substantiated by a significant coefficient $\hat{\beta} \approx 1.26 \times 10^{-5}$ in $V_t^{\asymp}/V_t = \alpha + \beta t + \varepsilon_t$, where V_t denotes the daily volume on the day-ahead market, which implies that the size of the imbalance market has also increased relative to the day-ahead market. Imbalance prices also still suffer from heavy shocks, in some instances resulting in negative prices.

On average, Figure 5b exhibits a small upward trend as well. However, it is not clear that the daily standard deviation is still increasing, as the process seems to have become more stable from 2014 onward. At least, we can conclude that technological developments have not led to more stable imbalance prices over the last ten years, but that there is no continuous increase in volatility either.

Quantitative metrics are included in Table 4 for both the daily price volatility and aggregated market volume. Similar to the day-ahead analysis, weekends typically exhibit relatively low price volatility and low trading volumes, accompanied by a low standard deviation. On Mondays, the imbalance market is most active with an average volume of 10,646 megawatt hour and an average price volatility of approximately €56. Panel B classifies daily price volatility and market volume by month of the year. Volatility and volume seem to be correlated in this regard, high price volatility typically coincides with large volumes. As observed with day-ahead prices, the imbalance market is most active during winter and fall. At least, this analysis shows that the imbalance market suffers from time fixed effects that cannot be ignored.

Subsample	Observations	Price volatility				Market volume in GWH[1]			
		Mean	Stdev	Min	Max	Mean	Stdev	Min	Max
Panel A: Per day of the week									
Monday	522	56.177	25.544	8.067	186.569	10.646	3.204	5.076	29.572
Tuesday	521	53.011	27.876	11.532	194.191	10.575	3.387	5.458	39.259
Wednesday	521	50.825	25.474	8.344	153.788	10.520	3.286	5.488	27.675
Thursday	522	53.646	26.981	6.453	155.888	10.506	3.220	5.430	24.629
Friday	522	50.796	25.737	7.521	160.222	10.431	4.340	4.966	81.480
Saturday	522	36.369	20.047	5.757	120.845	9.596	2.900	4.810	24.318
Sunday	522	39.567	21.134	8.213	159.080	9.574	2.909	4.259	22.690
Panel B: Per month									
January	310	54.205	26.224	7.791	139.395	11.134	3.173	5.607	29.572
February	282	46.555	24.551	9.388	160.222	10.064	2.527	5.315	19.153
March	310	50.713	27.537	8.301	194.191	10.763	3.254	6.158	27.675
April	300	45.015	25.049	6.453	193.773	10.253	3.499	5.687	24.309
May	310	44.056	24.233	7.540	159.080	9.872	3.675	5.076	39.259
June	300	46.594	25.574	8.213	163.645	9.855	5.195	4.259	81.480
July	310	47.951	27.266	5.757	154.633	10.033	2.974	4.810	22.803
August	310	45.663	23.157	8.308	127.402	9.839	3.055	4.587	20.110
September	300	49.592	25.297	9.276	142.701	9.747	2.935	5.130	23.852
October	310	49.273	26.607	7.521	165.304	10.195	3.263	5.076	26.768
November	300	52.586	25.228	11.283	153.788	10.183	2.922	5.562	24.629
December	310	51.092	26.487	11.244	176.478	11.175	2.942	6.369	22.690
Aggregated	3652	48.625	25.783	5.757	194.191	10.264	3.376	4.259	81.480

[1] Although the analysis is based on megawatt hours, the descriptive statistics are reported in 1,000 megawatt hours (gigawatt hours) for convenience.

Table 4: Descriptive statistics of imbalance prices

4 Methodology

4.1 Price predicting

In modelling and forecasting electricity spot prices, several specifications are employed. This section contains a detailed description of how each is composed. Every specification aims to provide a prediction for the electricity spot price P for observation (day) $t+1$ while assuming all information up to t to be available. Formally, let $f_m(t)$ denote the function representing the electricity spot price forecast for day t using model specification m. Then, for the $t+1$ forecast,

$$f_m(t+1) \equiv E_m(P_{t+1}|\Omega_t), \tag{4}$$

where the predicted value $E_m(P_{t+1})$ is estimated by model m, requiring that all prior price information up until t is known.

To adhere to the given requirement and to enable autocorrelation modelling, a rolling-window approach is employed, where a window of prior observations $P_{t-k+1}, P_{t-k+2}, \cdots, P_{t-1}, P_t$ with length k, satisfying $t-k+1 \geq 1$ and $t < T$ is used to estimate the models' coefficients with. These coefficients are then used to provide a forecast for P_{t+1}.

In this study, the models are built according to a bottom-up approach, where small and general models are subsequently refined by adding additional coefficients and levels of complexity to identify the optimal specification. Since the goal of the study is to optimize forecasts through time series analyses, the autocovariance structure is leading. Therefore, the first step is to maximize the explanatory power for electricity prices by finding the optimal AR(p) and, optionally, MA(q) parameters. Thereafter, electricity prices are forecast by employing different adjustments and extensions to the optimal model to get a better fit and to evaluate the hypotheses as defined in Chapter 2.

4.1.1 Models

The general approach in modelling time series is the AR(F)IMA model family, which is expressed as

$$P_t = \alpha + (1-L)^{-d}\vartheta(L)^{-1}\Theta(L)\varepsilon_t, \tag{5}$$

where L is the backshift operator with $L^i P_t = P_{t-i}$, d represents the difference parameter, and $\vartheta(L) = 1 - \vartheta_1 L - \vartheta_2 L^2 - \cdots - \vartheta_p L^p$ and $\Theta(L) = 1 + \Theta_1 L + \Theta_2 L^2 + \cdots + \Theta_q L^q$ denote the AR(p)

and MA(q) components, respectively (Baillie, 1996). The constant is stated α and the residuals ε_t satisfying $E(\varepsilon_t) = 0$.

If $p > 0$, the model contains p autoregressive lags and encompasses an AR(p) component. Parameter q similarly represents moving average lags—MA(q). If $d \neq 0$ and $d \in \mathbb{Z}$, the series is differenced d times in the estimation process and transformed from an ARMA model into an ARIMA specification. In the case when $d \neq 0$ but is not an integer, that is $d \notin \mathbb{Z}$, the model is classified as an ARFIMA model.

AR(p)

The starting point of modelling a time series with the presence of autocorrelation is a basic AR(p) specification. It is especially applicable when the data exhibits mean-reversion (Cuaresma et al., 2004). Essentially, the dependent variable, P_t, is expressed as a function of p of its own previous values. Formally, it is defined as a special case of Equation 5 with $d = 0$ and $\Theta(L) = 1$.

From the autocorrelation analysis in the data section it follows that partial autocorrelation is highly present up to the ninth lag. To confirm this observation, information criteria are constructed to assess the most plausible parameter p for a predictive context. Information criteria establish a balance between the combined fit of a set of coefficients and the number of coefficients in a model (Brooks, 2014, pp. 275-286). A low criterion implies that a model is more likely to deliver better forecasts. In empirical research, the Akaike (1974) and Bayesian (Schwarz, 1978) criteria are broadly employed and defined

$$AIC = \ln(\hat{\sigma}^2) + 2n/T \tag{6}$$

and

$$BIC = \ln(\hat{\sigma}^2) + \ln(T)n/T, \tag{7}$$

respectively. Here, the sample size is expressed by T and the number of coefficients in the model by n. Factor $\hat{\sigma}^2$ denotes the residual variance.

Since there is no consensus as to what criterion performs best, both the AIC and BIC are computed for AR(p) models with $p = 1, \cdots, 10$. Longer-horizon autocorrelation is likely to be captured by heterogeneous parameters and therefore $p = 10$ is assumed as the maximum lag. Furthermore, it is likely that partial autocorrelation spikes for lags > 10 are captured by time fixed effects, such as the day-of-the-week effect. By analyzing the entire data set in this regard, continuity of the autocovariance structure is implicitly assumed spanning both the in-sample and out-of-sample subsets, even though this may not be the case. To prevent violation of the

independence assumption for the out-of-sample range, information criteria are documented for both the entire sample and the in-sample subset, where the latter is leading.

Table 13 in Appendix B documents the results. Both the AIC and BIC are evidently lowest for the AR(9) model, consistent for both sets. Therefore, the AR(9) specification is assumed optimal and utilized as the fundamental model in the remainder of this study. However, the one-lagged alternative is maintained as well, mainly to provide insight in the power of heterogeneous parameters representing long-memory effects, similarly to Raviv et al. (2015).

ARMA(p,q)

Most time series do not follow straight autoregressive processes. Whereas AR(p) parameters capture autocorrelation directly with lags $1, \cdots, p$, moving average (MA) factors model the residuals of preceding observations. Combining both processes may fit the data better and is referred to as ARMA(p,q) specification. It is expressed using $d = 0$ in the fundamental model of Equation 5.

Series that follow an ARMA(p,q) structure typically exhibit autocorrelation and partial autocorrelation functions that both decline geometrically (Brooks, 2014, pp. 268-269). The former was clearly observed in the data section, and the latter to a lesser extent. Although computation processes become much more complex, adding a few moving average parameters may improve the predictive power. Information criteria are computed for ARMA(p,q) models for $p = 1, 9$ and $q = 1, 2, 3$ in Table 13 (Appendix B). First, adding 1, 2 and 3 moving average terms seems attractive for AR(1) models. Also, the AIC for the ARMA(9,2) model is smaller than the AIC for the basic AR(9) model, but the difference is only marginal. Furthermore, the BIC of this model exceeds that of the basic AR(9) model. Therefore, moving average terms will only be added as an extension later, not in the basic model.

ARFIMA(p,d,q)

As elaborated on in the data section, the Dutch electricity spot prices exhibit a fractionally integrated process due to the long-memory feature. The ARFIMA(p,d,q) specification essentially provides a bridge between modelling $I(0)$ and $I(1)$ series and it is used as a prediction instrument in several related studies (Gianfreda & Grossi, 2012; Haldrup & Ørregaard Nielsen, 2006). ARFIMA models essentially capture short-run and long-run effects along different dimensions, and are specified as in Equation 5, with a particular structure on d. Short-run influences are not differenced, $d = 0$, and long-run are differenced with $\hat{d}$, where the latter is estimated by a maximum likelihood optimization.

Table 13 in Appendix B depicts the information criteria for both an ARFIMA(1,d,0) and an ARFIMA(9,d,0) specification estimated on the entire data set, as well as on the in-sample range only. Both criteria in both samples confirm that ARFIMA frameworks fit the data slightly better than general AR models. An ARFIMA(9,d,2) specification performs best. In light of the third hypothesis, therefore, ARFIMA-type models will be estimated to correct for the long-memory feature in the data.

H($b_1, \cdots, b_n$)ARMA(p,q)

The goal of ARFIMA modelling is to capture high-order autocorrelation described as a long-memory property. However, estimating ARFIMA models is typically problematic with a high number of parameters due to its high complexity. Raviv et al. (2015) introduce the heterogeneous autoregressive (HAR) specification (Corsi, 2009) as a suitable alternative in this regard. In this specification, high-order autocorrelation is captured whilst maintaining a small number of coefficients and hence less complexity. Instead of expressing the autocorrelation up to order b by b coefficients, one coefficient is added representing the average of the previous b observations.

In terms of the fundamental model as defined in Equation 5, HAR parameters impose a specific structure on the autoregressive specification $\vartheta(L)$. That is

$$\vartheta(L) = 1 - \vartheta_{\text{AR},1}L^1 - \cdots - \vartheta_{\text{AR},p}L^p - \vartheta_{\text{HAR},b_1}\bar{L}^{1,b_1} - \cdots - \vartheta_{\text{HAR},b_n}\bar{L}^{1,b_n}, \tag{8}$$

where $\bar{L}^{1,b} = b^{-1}\sum_{i=1}^{b} L^i$ and denotes the average of the b previous lags.

Essentially, n coefficients are added representing different dimensions of aggregated autocorrelation. Following Raviv et al. (2015), two parameters are selected by default: weekly and monthly, that is H$(7, 30)$. For any autoregressive model with seven lags or more, the weekly factor will be omitted since the dependency of the previous week is already factored in in the autocovariance structure. The heterogeneous parameters are imposed as an alternative to the ARFIMA model for the third hypothesis.

4.1.2 Time fixed effects

Complementary to autocorrelation patterns, electricity spot prices suffer from calendar effects. Most importantly, prices differ between the days of the week. Similarly to Cuaresma et al. (2004), the fundamental model specification (Equation 5) is supplemented with the factor $\Upsilon_t = \sum_{i=2}^{7}\beta_{\Upsilon,i}I_{\Upsilon,t,i}$, resulting in

$$P_t = \alpha + (1-L)^{-d}\vartheta(L)^{-1}\Theta(L)\varepsilon_t + \sum_{i=2}^{7}\beta_{\Upsilon,i}I_{\Upsilon,t,i}, \tag{9}$$

where $\beta_{\Upsilon,i}$ denotes the coefficient belonging to the dummy variable $I_{\Upsilon,t,i}$ defined by the following system:

$$I_{\Upsilon,t,i} = \begin{cases} 1, & \text{if the day-of-the-week of } t = i \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $i = 1, 2, \cdots, 7$ and represents an array ranging from Monday (1) to Sunday (7). The factor $I_{\Upsilon,t,1}$—corresponding to Monday—is excluded to avoid a dummy multicollinearity.

In addition to day-of-the-week effects, prices typically differ between months. Similar to the day-of-the-week effect, the AR(F)IMA-type models are complemented with the factor $\Phi_t = \sum_{i=2}^{12} \beta_{\Phi,i} I_{\Phi,t,i}$:

$$P_t = \alpha + (1-L)^{-d} \vartheta(L)^{-1} \Theta(L) \varepsilon_t + \sum_{i=2}^{12} \beta_{\Phi,i} I_{\Phi,t,i}, \tag{11}$$

where $i = 1, 2, \cdots, 12$ and denotes the number of the month and $I_{\Phi,t,i}$ the dummy variable being equal to 1 if the month corresponding to t is equal to i, and 0 otherwise.

4.1.3 Spikes

Another important property of the spot prices is the presence of spikes, where extreme prices occur regularly and typically revert back to the local mean rapidly. In time series analyses, these spikes are problematic as their behavior is simply assumed to be normal. Especially when modelling autoregressive specifications, past spikes influence subsequent forecasts and result in prediction errors.

The first step in correcting for these spikes is identifying at which observations they occur. In equity markets, extreme prices are typically referred to as jumps, thereby relaxing the assumption that a rapid mean reversion process is present. The processes driving jumps are extensively researched in equity markets. Lee and Mykland (2007) introduce a non-parametric test to detect jumps in high-frequency stock markets, comparing the return from t to $t+1$ with the returns made in k previous observations, all based on logarithmic transformations. Andersen, Bollerslev, and Dobrev (2007) propose a similar methodology, evaluating individual returns with a constructed threshold. However, electricity markets fundamentally differ from stock markets. Most importantly, whereas stock returns are typically assumed to be completely stochastic, electricity prices suffer from predetermined patterns, making the identification of spikes substantially more difficult.

In electricity markets, different approaches have been proposed. Boogert and Dupont (2008) determine a fixed threshold for normal prices, where violations are identified as spikes. Alter-

natively, some studies incorporate a variable threshold as described by Clewlow and Strickland (2000), where the aggregated standard deviation is, on a recursive basis, used as an upper and lower bound for a short window of hourly power prices. However, for longer windows of prices, with lower frequencies—as is the case in the present study—these methods are likely to become problematic. Employing a fixed threshold fails to account for time variation and a variable threshold based on the standard deviation of the whole sample suffers from the same problem. Local spikes are unlikely to be accurately identified. Furthermore, many studies fail to correct for the seasonality component (Janczura et al., 2013).

In this study, a new framework is proposed as an attempt to improve the predictive power by filtering spikes, combining the seasonality property as analyzed by Janczura et al. (2013), the iterative approach of Lee and Mykland (2007) and the threshold method of Clewlow and Strickland (2000). To accurately identify price spikes, the unexpected component should be extracted from the price. Formally, prices are driven by a stochastic component z_t and a seasonality component y_t (Janczura et al., 2013). Assuming the seasonality factor consists of the day-of-the-week effect, the month-of-the-year effect and a (long) memory component in two dimensions (weekly and monthly), y_t it can be expressed as:

$$y_t = \alpha + \sum_{i=2}^{7} \beta_{\Upsilon,i} I_{\Upsilon,t,i} + \sum_{i=2}^{12} \beta_{\Phi,i} I_{\Phi,t,i} + \beta_{\Psi,7} P_{t-1,7} + \beta_{\Psi,30} P_{t-1,30} \tag{12}$$

where $I_{\Upsilon,t,i}$ and $I_{\Phi,t,i}$ denote day-of-the-week and month-of-the-year dummies respectively, α denotes a static constant and heterogeneous parameters $P_{t-1,b}$ are calculated as $b^{-1} \sum_{i=1}^{b} P_{t-i}$. Then, the model $P_t = y_t + z_t$ can be predicted to obtain an estimate of the seasonal component $\hat{y}_t = \hat{P}_t$ and the stochastic component $z_t \sim N(0, \sigma_z^2)$ by capturing the residuals. That is

$$z_t = P_t - \hat{y}_t. \tag{13}$$

The regression estimation process is based on the in-sample data range, to prevent data mining issues. One problem with this approach is that logarithmic transformations that are proposed in different studies are no longer available, since z_t can become negative as well. Although the variance structure would become more stable when a logarithmic transformation is applied, the benefit of deseasonalizing the prices is of higher importance.

In addition to average prices, price volatility differs through time. Therefore, the interval in which prices can be interpreted as normal is time varying. In line with Lee and Mykland (2007), the price series will be processed in an iterative rolling window algorithm, where prices that are identified as a spike are subsequently filtered and considered as such in follow-up iterations.

In this pre-filtering approach, stochastic prices z_t that exceed a certain confidence interval spanning the distribution around the previous k observations are identified as a spike. One would typically remove these extreme observations (outliers) from the sample in conventional statistical frameworks. However, since autocorrelation is key in this study, observations cannot simply be removed. Instead, individual spikes should be substituted by more reasonable alternatives. Weron (2006, p. 126) analyses three distinct methods to normalize spikes. Firstly, the observation could be limited to the boundary that it exceeds. Secondly, a dampening scheme can be applied, where the extreme value is decreased but only proportionally to the violation. The advantage of this approach is that extreme prices are retained to a certain extent. Thirdly, the price can be substituted by a similar price, for example that of the previous day. Janczura et al. (2013) add another method when deseasonalized data is employed, where the spike component z_t is simply substituted by the mean of its corresponding rolling window $\bar{z}_{t,k}$ (averaging). Although the window-specific means are on average equal to zero as forced by the deseasonalization model, time variation would be neutralized if the averaging method would simply replace z_t by zero.

Empirical research has not yet reached consensus as to an optimal method to deal with spikes. Therefore, all four proposed methods (limiting, dampening, replacing and averaging) are implemented. The dampening approach is described by a formula incorporating a logarithmic transformation (Shahidehpour et al., 2002, p. 83). However, deseasonalized data also yields negative values that cannot be converted to logarithms. To resolve that restriction, the dampening scheme will be applied on the log-normal boundary prices, derived back from the residual analysis. In other words, while the regular identification process is based on residuals, the dampening technique is applied on the boundary prices to prevent negative inputs.

The pre-filter framework is documented in the following algorithm, requiring two static inputs, k and z^*. A graphical analysis showed that a rolling window size of $k = 30$ is an acceptable assumption, since it accurately identifies most of the observations at which a spike is expected. Increasing the size does not add additional power and lowering k causes unstable and unreliable confidence intervals. For the identification of spikes using the variable threshold approach, literature tends to select $z^* = 3$ (Clewlow & Strickland, 2000; Keles et al., 2012). This critical value seems reasonable in the current data set as well, based on a graphical investigation.

Algorithm 1 *(Spike pre-filter). Let k denote the rolling window size, P_t the observed price at time t, $\hat{y}_t$ the seasonal price component at time t, T the number of observations for the stochastic electricity prices z_t, z^* the critical value representing the confidence interval and Ξ_t a categorical variable $\in \{1, -1, 0\}$ classifying observation t as a positive spike, negative spike or no spike,*

respectively. Then, for each $t = k+1, \cdots, T$*:*

1. *Calculate mean of* z_t *over* k *previous observations. That is* $\bar{z}_{t,k} = k^{-1} \sum_{j=1}^{k} z_{t-j}$.
2. *Similarly, compute the standard deviation by* $\sigma_{t,k} = \sqrt{(k-1)^{-1} \sum_{j=1}^{k} (z_{t-j} - \bar{z}_{t,k})^2}$.
3. *Assign the spike factor based on the boundaries* $\Xi_t = \begin{cases} 1, & \text{if } z_t > z_t^{\{1\}} = \overline{z}_{t,k} + z^* \sigma_{t,k} \\ -1, & \text{if } z_t < z_t^{\{-1\}} = \overline{z}_{t,k} - z^* \sigma_{t,k} \\ 0, & \text{otherwise.} \end{cases}$
4. *Derive upper price bound* $P_t^{\{1\}} = \hat{y}_t + z_t^{\{1\}}$ *and lower price bound* $P_t^{\{-1\}} = \hat{y}_t + z_t^{\{-1\}}$.
5. *Only if* $\Xi_t \neq 0$*, apply treatment scheme on* P_t *conditional on selected method:*

$$P_t = \begin{cases} \hat{y}_t + z_t^{\{\Xi_t\}}, & \text{with limiting scheme} \\ P_t^{\{\Xi_t\}} + P_t^{\{\Xi_t\}} \log(P_t / P_t^{\{\Xi_t\}}), & \text{with dampening scheme} \\ \hat{y}_t + z_{t-1}, & \text{with replacing scheme} \\ \hat{y}_t + \bar{z}_{t,k}, & \text{with averaging method.} \end{cases}$$

6. *Enter back the stochastic component* z_t *for the purpose of calculating the statistics for the next iterations. That is* $z_t = P_t - \hat{y}_t$.

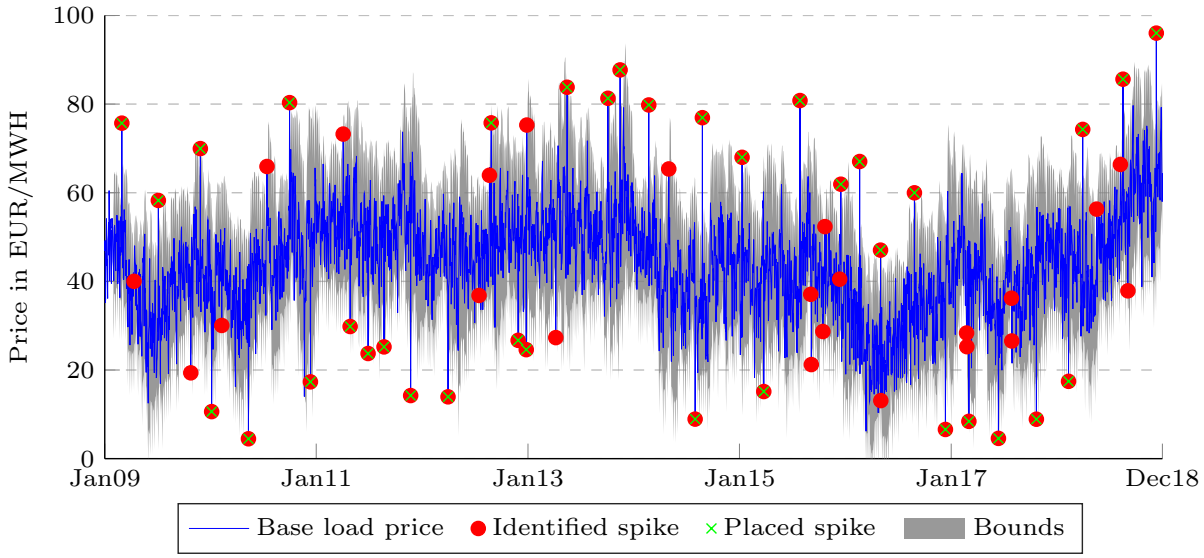
A major advantage of this iterative approach is that it is only dependent on preceding observations. Therefore, for any new observation that enters the sample, this procedure can be repeated only for that observation, making it a convenient algorithm to work with in practice.

Even though the pre-filter algorithm is based on actual observed dynamics, the fact that it is not used before in this particular composition requires additional validation. One way to assess its reliability is to apply it on a fictive price series, composed by a specific data generating process (DGP). In this study, the simulation setup is based on Janczura et al. (2013) to generate electricity prices from 2008 up to 2019. The algebraic specifics are highlighted in Appendix C. Subsequently, price spikes are added to the series by splitting the sample of 3,652 prices in pieces of a 100 observations, excluding the first 52 for convenience, which is justified since the pre-filter algorithm starts after the fist rolling window anyway. In each of the 36 sub-samples, one random spike (positive or negative) is added by replacing one random observation with $\bar{P} + \delta\lambda\sigma$, where $\bar{P}$ and σ denote the average electricity price and standard deviation of electricity prices in that specific sub-sample, respectively. Factor δ represents the sign of the spike, $\delta \in \{-1, 1\}$, whereas λ expresses the spike's magnitude, which is randomly picked and satisfies $3.5 \leq \lambda \leq 4.5$.

Figure 6 presents the development of a simulated series, in conjunction with identified spikes using the limiting approach. It is evident that the spike filtering algorithm has detected 100% (36 out of 36) of the spikes that were randomly inserted. On top of that, it identified 23 prices that are not classified as spikes. It follows that the limiting pre-filter algorithm performs well in

this case in terms of power, but the size property requires additional attention.

Figure 6: Simulation of a price series with inserted spikes



Notes: In this figure, the inserted spikes (36) are presented as a green cross within a simulated series (see Appendix C) of daily base load electricity prices ranging from January 2009 to December 2018. Identified spikes (59) are presented as a red dot. The identification is based on the iterative rolling window algorithm (Algorithm 1) using the limiting scheme. The grey area represents the confidence interval for spike detection.

Method	Power			Size		
	Average score	$\geq 90\%$	$\geq 95\%$	Average score	$\leq 1\%$	$\leq 5\%$
Limiting	96.33%	95.20%	62.60%	0.67%	98.70%	100.00%
Dampening	95.94%	95.30%	56.60%	0.60%	99.60%	100.00%
Replacing	86.39%	25.10%	3.60%	0.46%	100.00%	100.00%
Averaging	86.62%	27.10%	4.30%	0.46%	100.00%	100.00%

Notes: This table contains results of the spike pre-filter algorithm, applied on 1,000 simulations of each treatment method. The power of the algorithm represents the ability to detect spikes that are present and is calculated as the fraction of randomly inserted spikes that were identified to the total number of randomly identified spikes. A score of 100% implies that the algorithm works perfectly in identifying the spikes that were inserted. The filter's statistical size is calculated as the number of identified spikes that are not real spikes divided by the total number of observations that do not exhibit a spike, that is $3,652 - 36 = 3,616$. Next to the average score of both metrics, the relative number of simulations that performed better than or equal to a certain threshold are reported.

Table 5: Simulation results of the spike pre-filter algorithm

To corroborate this single simulation, another 1,000 simulations have been generated for all four pre-filter methods. For every simulation, both the power and size are computed representing the ability to identify present spikes and the probability that the algorithm identifies a spike that is not a real spike, respectively. The results are summarized in Table 5. A high power and low size imply that the algorithm is a sound technique to detect price spikes. In terms of size, every method seems appropriate, none identifies more than 5% of non-spikes as a spike. It is evident that the limiting and dampening methods are powerful, on average both identifying approximately 96% of the inserted spikes. Furthermore, approximately 95% of all simulations

produced a power score with a minimum of 90%. However, the replacing and averaging methods perform substantially worse, both having an average score close to 86%, and an extremely small fraction that identifies at least 90% of the spikes.

Evidence suggests that the pre-filter algorithm works well in identifying the spikes, but only when the limiting or dampening method is applied. It should be noted though that some components in the simulation setup are relatively arbitrary, such as the fixed sub-sample structure and the predetermined magnitudes of the spikes. However, since the simulated price processes are strongly in line with observed EPEX spot prices, the simulation analysis is considered reliable.

4.1.4 Optimization

Although the theoretical hypotheses focus on relatively standard AR(p) models with, optionally, HAR components, it may be possible to further improve these by including moving average terms or allowing models that can handle different distributions. Although no formal hypotheses have been formed to substantiate possible improvements, it is still useful to assess whether technical adjustments can decrease prediction errors.

Firstly, moving average (MA) terms are added. As documented in the data section, adding two MA terms may improve forecasting accuracy, although the improvement is expected to be marginal, since information criteria are not unambiguously lower.

Secondly, it is evident that the distribution of electricity prices cannot be assumed normal, due to heavy tails and asymmetry. Although the prices are sufficiently covariance stationary, prediction models are sensitive for atypical distributions. In that regard, a log-normal transformation, as employed by Cuaresma et al. (2004), may reduce the variance and generate a more stable distribution. The optimal model(s) as determined by the first four hypotheses will therefore be estimated on $\log(P_t)$ as well, where the predicted log-normal prices are subsequently converted back to normal prices. By transforming normal models with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ to natural logarithms, it is implicitly and incorrectly assumed that $\exp(\varepsilon_t) \sim \log N(0, \sigma_\varepsilon^2)$. Consequently, forecasts should be adjusted to correct for the incorrectly specified mean (Mount et al., 2006). Algebraically, the forecast function becomes

$$f_m(t+1) \equiv \exp\left[E_m(P_{t+1}|\Omega_t) + \frac{1}{2}\hat{\sigma}_\varepsilon^2\right], \tag{14}$$

where $E_m(P_{t+1}|\Omega_t)$ denotes the prediction for time $t+1$ based on log-normal model m assuming that all pricing information up until t is known and $\hat{\sigma}_\varepsilon^2$ the estimated variance of the forecast errors.

4.1.5 Evaluation

Different measures are available to assess the accuracy of forecasts against realized observations. Since the sign of the deviation is typically irrelevant, most measures examine absolute values of deviations from the real observations. The most direct measure is the mean average error (MAE), which captures the average absolute error over the predicted values. In the finite sample of electricity prices $1, \cdots, T$, let $t_0, \cdots, t_1$ denote the prediction window with $1 < t_0 \leq t_1 \leq T$ and length $t_1 - t_0 + 1$ that is forecast using model m. Then, the MAE function is defined by

$$MAE_m(t_0, t_1) \equiv \frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} \big|P_t - f_m(t)\big|, \tag{15}$$

where P_t and $f_m(t)$ denote the real and forecast price at time t, respectively.

A similar alternative is the mean squared error (MSE), which has the additional property that it is measured quadratically, thereby punishing extreme errors more aggressively. This metric is more appropriate when large deviations are disproportionally undesirable. It is defined as

$$MSE_m(t_0, t_1) \equiv \frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} \big[P_t - f_m(t)\big]^2. \tag{16}$$

These two metrics however tell little about the relative deviation of the predicted prices—they only report absolute values. The mean average percentage error (MAPE) provides the average error in terms of a percentage of real prices. It is calculated as

$$MAPE_m(t_0, t_1) \equiv \frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} \left|\frac{P_t - f_m(t)}{P_t}\right|, \tag{17}$$

where $f_m(t)$ and P_t denote the forecast function for model m and real price and day t, respectively.

In comparing price prediction models, a lower error measure indicates better forecasting performance. However, the extent to which a certain model outperforms another model, at least for assessing the validity of the stated hypotheses, should be assessed by means of a statistical test. A too small difference between two models could simply be the result of idiosyncratic factors, for example a sample selection bias. Diebold and Mariano (1995) propose a pairwise statistical test to compare forecast performances of two models based on their loss functions. Their procedure involves standardizing the differences between two series of loss functions to their distribution in a single number. Mainly due to its simplicity and its universal applicability, many studies in electricity prices use this test as a reliable instrument to evaluate forecasts with

(Cuaresma et al., 2004; Lago, Ridder, Vrancx, & Schutter, 2018; Nowotarski & Weron, 2016). For that reason, the Diebold and Mariano (1995) procedure is employed in the present study as well to test whether a more sophisticated model m_2 outperforms its basic alternative m_1.

Let $\ell_m^{\{x\}}$ denote the x-based loss function for observation t with $x \in \{AE, SE, APE\}$, corresponding to observation-specific metrics for the MAE, the MSE and MAPE, respectively. Formally,

$$\ell_m^{\{x\}} \equiv \begin{cases} |P_t - f_m(t)|, & \text{if } x = AE, \\ [P_t - f_m(t)]^2, & \text{if } x = SE, \\ P_t^{-1}|P_t - f_m(t)|, & \text{if } x = APE. \end{cases} \tag{18}$$

Then, the loss differential given x between model m_1 and m_2 is calculated as

$$d_{m_1,m_2}(t)|x = \ell_{m_1}^{\{x\}}(t) - \ell_{m_2}^{\{x\}}(t). \tag{19}$$

Under the null hypothesis, the expected difference in predictive accuracy is equal to zero. That is

$$H_0 : \mathbb{E}[d_{m_1,m_2}(t)] = 0. \tag{20}$$

In this study, model comparisons are initially based on one-sided evaluations, since a refined model m_2 is expected to outperform its more basic alternative m_1. Therefore, the alternative hypothesis is defined as

$$H_a : \mathbb{E}[d_{m_1,m_2}(t)] > 0. \tag{21}$$

To evaluate whether the null hypothesis should be rejected, Diebold and Mariano (1995) prescribe the test statistic

$$t_{m_1,m_2}^{DM} = \frac{\bar{d}_{m_1,m_2}}{\sqrt{\hat{\sigma}_{m_1,m_2}^2/N}}, \tag{22}$$

where $\bar{d}_{m_1,m_2}$ and $\hat{\sigma}_{m_1,m_2}^2$ denote the mean and estimated unconditional variance of the performance differentials $d_{m_1,m_2}(t)$ of N point forecasts, respectively[1]. In estimating the variance, an estimator is employed that is referred to as a Newey-West type. Therefore, potential autocorrelation and heteroskedasticity issues are corrected for. Since m_1 and m_2 can simply be flipped, outperformance can be approached from two directions. Based on the student t distribution,

[1]The Diebold and Mariano (1995) test statistics are computed using MATLAB.

it follows that the null hypothesis should be rejected if $t^{DM}_{m_1,m_2} > t^*_{5\%} = 1.645$. Conversely, $t^{DM}_{m_1,m_2} < -1.645$ simply implies outperformance of model m_1 relative to m_2.

When comparing many models, however, pairwise procedures are insufficient as an instrument to identify the best performing specification(s). Typically, the data set does not yield a set of unambiguously superior models, due to a lack of complete information (Hansen, Lunde, & Nason, 2011). For that purpose specifically, Hansen et al. (2011) propose the model confidence set (MCS) procedure, which has been applied in electricity pricing (Bordignon, Bunn, Lisi, & Nan, 2013), that aims at identifying a set of models $\mathcal{M}^*_{1-\alpha}$ that are superior in an initial model set $\mathcal{M}^0$. Formally, the superior set of models $\mathcal{M}^*_{1-\alpha} \subseteq \mathcal{M}^0$ is defined by

$$\mathcal{M}^* \equiv \{i \in \mathcal{M}^0 : \mu_{i,j} \leq 0 \; \forall \; j \in \mathcal{M}^0\}, \tag{23}$$

which represents a selection of models that perform equally and outperform every other model on an $(1-\alpha)\%$ confidence level. Through a series of sequential equivalence tests between models $i, j \in \mathcal{M}^0$, inferior models are removed from the subset $\mathcal{M}$ until this process converges. The null hypotheses are similar to the Diebold and Mariano (1995) procedure,

$$H_{0,\mathcal{M}} : \mu_{i,j} = 0, \; \forall \; i, j \in \mathcal{M}, \tag{24}$$

where $\mu_{i,j} = \mathbb{E}[\bar{d}_{i,j}(t)]$ (see Equation 19). If $H_{0,\mathcal{M}}$ is rejected, $\delta_{\mathcal{M}} = 1$ and $\delta_{\mathcal{M}} = 0$ otherwise. To formally test these null hypotheses, pairwise student-t statistics are constructed similar to the Diebold and Mariano (1995) statistic. The p-value is then calculated for the highest absolute student-t statistic that represents the most extreme inter-model difference[2]. If the p-value is smaller than 5%, the null hypothesis of equal predictive performance is rejected and the worst performing model, according to elimination rule $e_{\mathcal{M}}$, is removed from subset $\mathcal{M}$. Formally, the Hansen et al. (2011) algorithm is described as follows.

Algorithm 2 *(Model confidence set procedure). Let $\mathcal{M}^0$ denote a set of predictive models.*

1. *Assign $\mathcal{M} = \mathcal{M}^0$.*
2. *Test whether $H_{0,\mathcal{M}}$ holds based on confidence level α.*
   - *If $\delta_{\mathcal{M}} = 1$, eliminate model $e_{\mathcal{M}}$ and rerun step 2.*
   - *If $\delta_{\mathcal{M}} = 0$, finish the algorithm and set $\hat{\mathcal{M}}^*_{1-\alpha} = \mathcal{M}$.*

4.2 Imbalance dynamics

The remainder of the methodology chapter focuses on the relation between the day-ahead spot prices and the dynamics in the imbalance market. In particular, the goal is to assess whether the

[2]The MCS procedures are performed using the MFE Toolbox by Kevin Sheppard in MATLAB.

predictability of Dutch electricity spot prices is related to quantities and prices on the imbalance market. To establish such a framework, the first step is to express the spot price predictability numerically.

4.2.1 Predictability measures

The predictability of spot prices can be approximated in different ways—no unique method exists. Electricity prices suffer from strong effects of autocorrelation. Therefore, the extent to which the price at $t+1$ is predictable can be simply measured by the correlation between the hourly prices at t and hourly prices at $t+1$. The first and most direct measure of predictability at day t is

$$\rho(t) \equiv \frac{\text{Cov}(\boldsymbol{P}_{t-1}, \boldsymbol{P}_t)}{\sigma(\boldsymbol{P}_{t-1})\sigma(\boldsymbol{P}_t)}, \tag{25}$$

where $-1 \leq \rho(t) \leq 1$ and $\boldsymbol{P}_t$ denotes the 1×24 vector $[P_{t,1}, P_{t,2}, \cdots, P_{t,24}]$ including all 24 hourly prices of day t and $\sigma(\cdot)$ the standard deviation of a price vector. Note that the autocorrelation metric is not perfect to assess predictability, since relatively high inter-day differences could be, in fact, be predictable on the basis of time fixed effects. However, these effects can be corrected for ex post by including control variables. Furthermore, it is crucial to take into account that hourly prices are in fact determined by the market system, whereas base load prices merely represent an aggregated price. Therefore, utilizing hourly prices in assessing the predictability is considered optimal, and this measure is expected to be the most direct and reliable.

A problem with measuring predictability by comparing two individual days, is that idiosyncratic factors can rapidly influence the results. Estimating autocorrelation on the base load price using a broader interval might give a more reliable estimate, although becoming less applicable to a specific day. The autocovariance function for lag $|v| < t_1 - t_0 + 1$ based on a price series $P_{t_0}, \cdots, P_{t_1}$ can be written as a function with three arguments. That is

$$\tau(v, t_0, t_1) \equiv \frac{1}{t_1 - t_0 + 1} \sum_{i=t_0}^{t_1-|v|} (P_i - \bar{P}_{t_0,t_1})(P_{i+v} - \bar{P}_{t_0,t_1}), \tag{26}$$

where $\bar{P}_{t_0,t_1} = (t_1 - t_0 + 1)^{-1} \sum_{i=t_0}^{t_1} P_i$ and represents the average electricity price in the window $t_0, \cdots, t_1$. Let $k = 20$ denote the length, in days, of the rolling window to estimate the $v = 1$ autocorrelation for each $t \geq k$. Then, the autocorrelation function from the perspective of t can be written as

$$\gamma(t) \equiv \frac{\tau(1, t-19, t)}{\tau(0, t-19, t)} = \frac{\sum_{i=t-19}^{t-1}(P_i - \bar{P}_{t-19,t})(P_{i+1} - \bar{P}_{t-19,t})}{\sum_{i=t-19}^{t}(P_i - \bar{P}_{t-19,t})^2}, \tag{27}$$

by construction satisfying $-1 \leq \gamma(t) \leq 1$. A high $\gamma(t)$ implies a strong degree of predictability.

Finally, model implied accuracy can be employed as an instrument for measuring predictability. More specifically, loss functions can be used as a proxy. Let $\ell(t)$ denote the absolute prediction error for day t given an optimal forecasting model m, which is to be determined in the results section. We have

$$\ell(t) \equiv \big| f_m(t) - P_t \big|, \tag{28}$$

where $f_m(t)$ denotes the forecasting function and P_t the observed price. It should be noted, however, that model implied predictability is subject to model risk. That is, the model that is selected could be misspecified and not realistically reflecting predictability in real markets. This measure is therefore assumed the least reliable.

4.2.2 Relation analysis

Hypotheses five and six aim towards finding (negative) relationships between the predictability of day-ahead spot prices and intra-day imbalance activity. First, it is expected that the intra-day price volatility in the imbalance market is negatively related to the predictability of day-ahead power prices. To assess the existence of this relationship, a regression analysis is deployed covering the whole sample. We define $g(t)$ as the generalized predictability function with $g \in \{\rho, \gamma, \ell\}$. Then, the regression model is expressed as

$$\sigma_t^{\asymp} = \alpha + \beta \cdot g(t) + \varepsilon_t, \tag{29}$$

where α denotes a static constant and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ the error term.

The sixth hypothesis focuses on the daily aggregated volume on the imbalance market, and expects a negative relationship with day-ahead predictability as well. Similarly, the model is

$$V_t^{\asymp} = \alpha + \beta \cdot g(t) + \varepsilon_t, \tag{30}$$

with constant α and error term $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Note that when the model implied predictability is analyzed, only the out-of-sample day-ahead range is included.

Although both models expect a negative relationship, it is also interesting to observe any positive betas. Correspondingly, the betas will be tested in a two-sided manner:

$$H_0 : \beta = 0$$
$$H_a : \beta \neq 0.$$

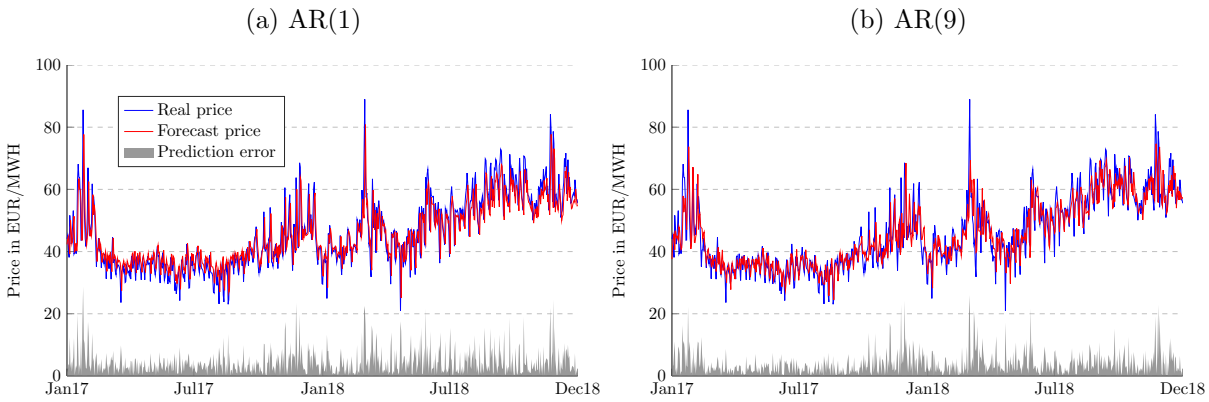
The null hypothesis is rejected if the test statistic $|t^{OLS}| = |\hat{\beta}/SE(\hat{\beta})| > t^*_{5\%/2} = 1.960$ (Brooks, 2014, pp. 99-103).

To prevent heteroskedasticity resulting in incorrect standard errors, both regressions incorporate robust White (1980) errors. Additionally, an apparent pitfall of these models is endogeneity—omitted variables could result in a spurious relationship. For example, the correlation between day-ahead prices is presumably high between Saturday and Sunday. But this pattern is present in imbalance markets as well: price volatility is very similar between these days. If this effect outweighs the other five days, then it is possible that the regression models yield significant coefficients. To limit the possibility of such a bias, various control variables are added to both regression models. Firstly, the models are complemented with day-of-the-week and month-of-the-year dummies. Secondly, the t variable is added to filter out the average trend.

5 Forecasting

Initially, basic AR(1) and AR(9) models are employed as forecasting instruments for the out-of-sample range. Figures 7a and 7b depict the graphical accuracy of these models as opposed to the real prices. Although the average absolute prediction error of the AR(9) model seems lower, it does not predict extreme price spikes well. In this regard, the AR(1) model seems to perform a substantially better. Further, it is evident that one or more fundamental components are missing—both models show structural deviations that could potentially be reduced by controlling for time fixed effects. Lastly, price spikes evidently yield the highest prediction errors and factoring these in might improve performance.

Figure 7: Basic autoregressive model forecasts

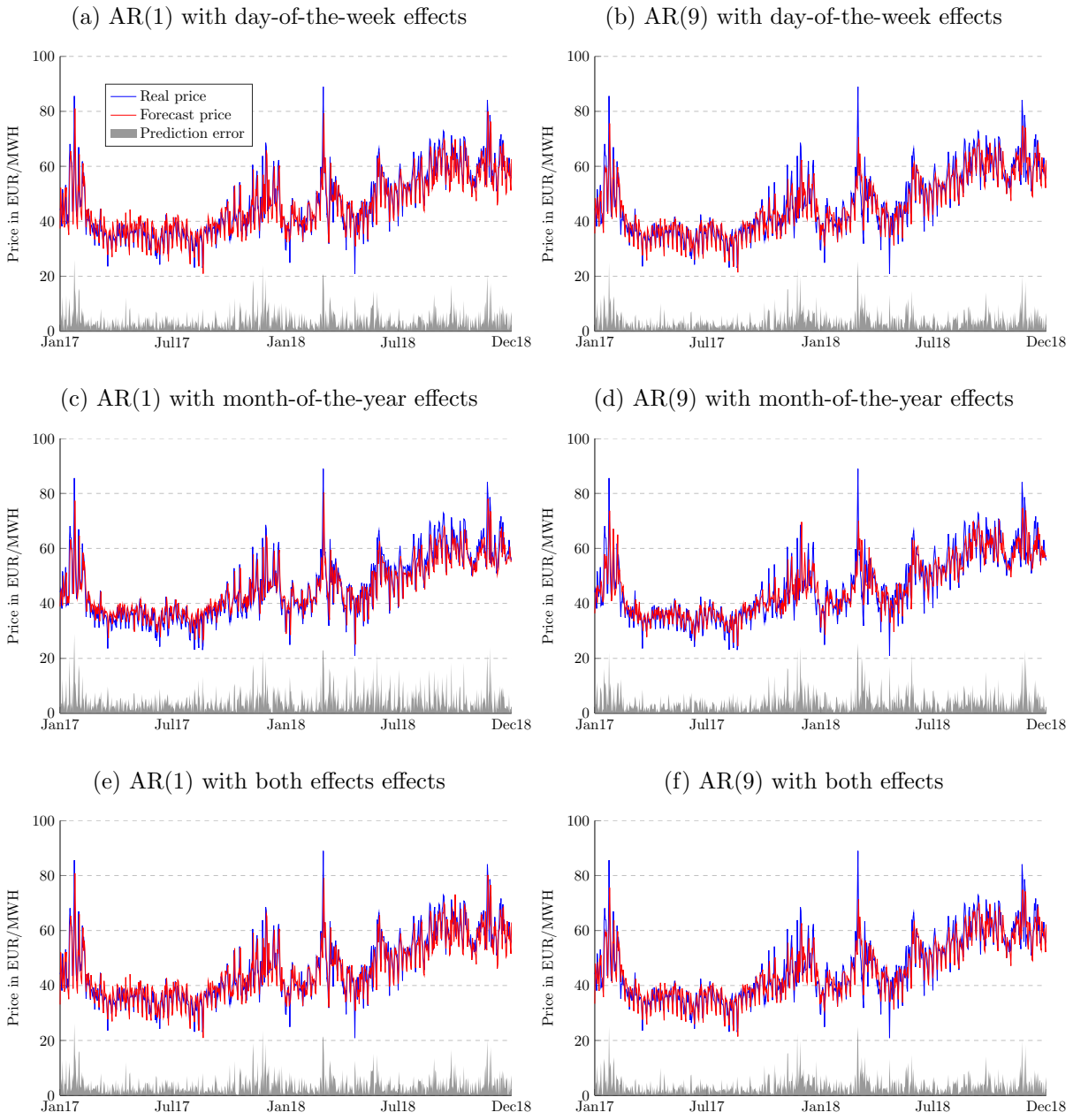


Notes: Figures (a) and (b) illustrate the predictive performance of a basic AR(1) and AR(9) model, respectively, relative to the observed out-of-sample electricity prices.

5.1 Time fixed effects

The first extension to the basic models is the day-of-the-week effect. Figures 8a and 8b illustrate the predictive performance of an AR(1) and AR(9), respectively, incorporating day-of-the-week dummies. At least, it is clear that the performances have improved as compared to the basic models without dummies. Again, it is unclear which autoregressive specification performs best in this regard, since the AR(9) model seems to underperform with extreme prices. Similarly, Figures 8c and 8d describe the price forecasts when a month-of-the-year effect is included. When compared to the basic autoregressive models, it is not clear whether this time fixed effect improves performance. Lastly, the extension with both the day-of-the-week and month-of-the-year effect is presented in Figures 8e and 8f. Graphically, it is inconclusive whether these specifications show superior performance as compared to the previous models.

Figure 8: Autoregressive model forecasts with time fixed effects.



Notes: The figures on the left (right) side exhibit the performance of electricity price forecasts for the out-of-sample range based on an AR(p) model with $p = 1$ ($p = 9$). Additionally, from top to bottom, the models are extended with day-of-the-week and month-of-the-year effects.

Table 6 contains the predictive performances of the AR(1) and AR(9) specifications complemented with time fixed effects. It is clear that all three loss functions (MAE, RMSE and MAPE) are inter consistent. That is, the conclusions inferred from this table are independent of the choice of the loss function and therefore robust in that regard.

Panel A presents the performance of the inclusion of time fixed effects in the AR(1) model. It follows that the day-of-the-week effect improves the performance significantly, as indicated by substantial lower loss functions and 1% significant Diebold and Mariano (1995) statistics.

Model	MAE	MAE vs basic	MSE	MSE vs basic	MAPE	MAPE vs basic
Panel A: AR(1)						
Basic	4.287		35.424		9.53%	
Day-of-week	3.689	5.37***	25.938	6.10***	8.04%	5.59***
Month-of-year	4.273	0.75	35.515	-0.36	9.48%	1.12
Both	3.687	5.32***	26.122	5.89***	8.03%	5.59***
Panel B: AR(9)						
Basic	3.841		28.842		8.38%	
Day-of-week	**3.503**	3.97***	**23.330**	4.65***	**7.63%**	3.79***
Month-of-year	3.863	-2.06**	29.014	-1.66**	8.43%	-1.97**
Both	**3.514**	3.79***	**23.510**	4.44***	**7.65%**	3.63***

Notes: This table documents predictive performance metrics for both AR(1) and AR(9) specifications. The mean average error (MAE), mean squared error (MSE) and mean average percentage error (MAPE) measures are reported. In addition, Diebold and Mariano (1995) test statistics are included comparing row and column specifications, where a positive (negative) value indicates that the row (column) model performs better either not significantly, or significantly on a 10% level, 5% level or 1% level, denoted by *, ** and ***, respectively. An underlined metric indicates that, according to this metric, the specified model is present in the 95% model confidence set within the panel. Metrics in **bold** imply presence in the 95% confidence set of all panels combined.

Table 6: Forecasting performance with time fixed effects

The day-of-the-week AR(1) specification yields forecasts that are, on average, approximately 1.5 percent point more accurate than the basic AR(1) model, resulting in a average prediction error of 8.04%. However, the month-of-the-year effect is not as meliorative. Adding a month-of-the-year factor only lowers the average prediction error with 0.05 percent points, confirmed by the insignificant DM statistic for all three metrics. Logically, combining both the day-of-the-week effect and the month-of-the-year effect yields similar prediction errors to the day-of-the-week effect only, confirmed by the presence of both models in the model confidence set. Since the day-of-the-week model consists of substantially less coefficients, this model is preferred. Consequently, in AR(1) forecasting models, incorporating a month-of-the-year effect is trivial and the day-of-the-week alternative is considered optimal.

For the AR(9) specifications, documented in Panel B, the results are similar. The day-of-the-week model significantly outperforms its basic alternative. Including a month-of-the-year factor only insignificantly improves the forecasting accuracy relative to the basic specification, and yields a slightly higher prediction error in conjunction with the day-of-the-week effect.

The first hypothesis was defined as follows:

Hypothesis 1 *Adding a day-of-the-week effect improves the forecasting performance for Dutch day-ahead electricity prices.*

From the analysis in this section, sufficient evidence is found to support this claim. Incorporating the day-of-the-week effect enhances the accuracy of the predictions.

Hypothesis 2 *Incorporating a month-of-the-year dependency improves statistical forecasting power for Dutch electricity spot prices.*

In contrast, no evidence was found to support the second hypothesis. Despite the monthly price differences as observed in the data section, accounting for this in forecasting models does not improve power. Consequently, the month-of-the-year effect will be omitted in further analyses.

Table 6 further documents that AR(9) specifications with day-of-the-week effects and both effects outperform AR(1) alternatives. Including nine lags is therefore considered optimal. However, this does not imply that AR(1) models should be removed from any further analyses. AR(1) specifications could very well outperform nine-lagged alternatives when long memory effects are considered. In conclusion, the day-of-the-week AR(9) model performs best and leads to an average prediction error of 7.63%.

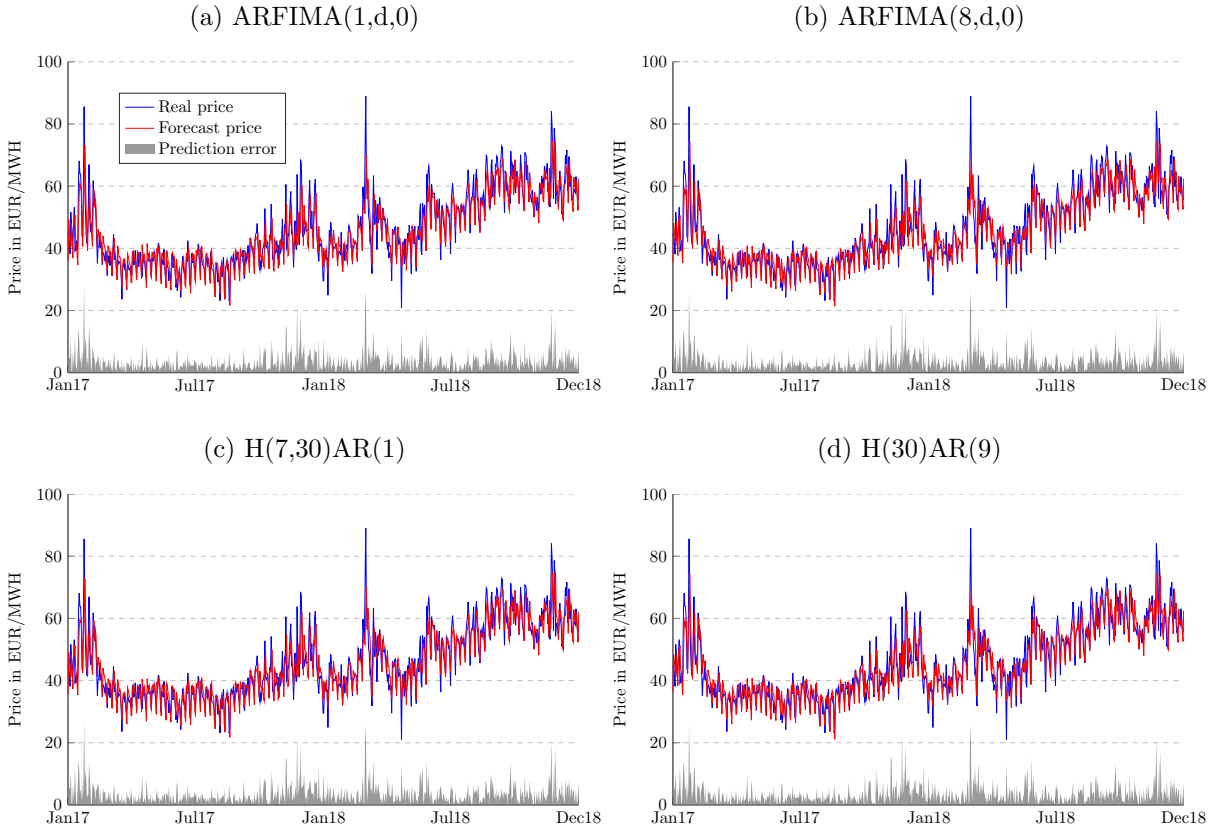
5.2 Long memory

In addition to calendar effects that have a fixed effect on the development of daily electricity prices, the pricing process tends to move over time due to the long memory property. To account for this, two model classes are introduced. Firstly, ARFIMA models are estimated. Day-of-the-week supplemented ARFIMA(9,d,q) models, however, are subject to severe unit circle errors, which implies that the data exhibits non-stationarity in some combinations, due to the high complexity. Therefore, ARFIMA(8,d,q) specifications are estimated instead. Secondly, heterogeneous parameters are introduced representing different dependency effects. The AR(1) model is complemented with two heterogeneous parameters representing a weekly and monthly subjection. For the AR(9) model, only a monthly parameter is added, as it already encompasses a weekly dependency effect by its first seven lags.

Figures 9a to 9d present graphical analyses on the electricity price forecasts incorporating long memory features. It is not unambiguous that accounting for this improves performance. Spikes are still subject to large prediction errors. However, the extensions may very well lead to lower average errors and thus better performance. A quantitative analysis will shed light thereon.

The performance measures are documented in Table 7. Panel A describes the predictive accuracy from the perspective of the AR(1) model with day-of-the-week effects. It is evident that including long memory components decreases forecasting errors significantly, consistent for every loss function. The MAPE is decreased from 8.04% to 7.59%, 7.68% and 7.59% in the ARFIMA(1,d,0), H(7)AR(1) and H(7,30)AR(1) models, respectively. Further, the model

Figure 9: Long-memory price forecasts



Notes: The figures on the left (right) side exhibit the performance of electricity price forecasts for the out-of-sample range based on an AR(p) model with $p = 1$ ($p = 9$ or, in the case of ARFIMA specifications, $p = 8$). All specifications are equipped with day-of-the-week fixed effects.

confidence set indicates that the H(7)AR(1) specification is outperformed by the former and the latter in terms of the MAE.

For the AR(9) models, the results are less consistent. ARFIMA(8,d,0) day-of-the-week significantly beat the AR(9) model in terms of the MAE and the MAPE, but not in terms of the MSE. The H(30)AR(9) model, however, does not yield any significant improvement on at least a 5% level. Adding a monthly dependency does parameter does not significantly improve the AR(9) specification. The panel-specific 95% model confidence set that includes all alternatives including the original AR(9) model, further implies that all alternatives perform relatively equally.

The hypothesis addressing the long memory property was stated as:

Hypothesis 3 *The forecasting accuracy of Dutch electricity spot prices improves by accounting for a long-memory feature.*

Based on the previous reasoning, sufficient evidence is found for the presence of a long memory effect and forecasting improvements due to modelling it, most importantly when ARFIMA

Model	MAE	MAE vs AR	MSE	MSE vs AR	MAPE	MAPE vs AR
Panel A: AR(1)						
AR(1)	3.689		25.938		8.04%	
ARFIMA(1,*d*,0)	**3.486**	3.88***	**23.317**	3.33***	**7.59%**	4.08***
H(7)AR(1)	**3.527**	2.68***	**23.668**	2.72***	**7.68%**	2.75***
H(7,30)AR(1)	**3.478**	3.49***	**23.203**	3.30***	**7.59%**	3.44***
Panel B: AR(9)						
AR(9)	**3.503**		**23.330**		**7.63%**	
ARFIMA(8,d,0)	**3.477**	1.72**	**23.242**	0.45	**7.55%**	2.46***
H(30)AR(9)	**3.478**	1.55*	**23.212**	0.63	**7.58%**	1.43*

Notes: This table describes the predictive performances in terms of the MAE, the MSE and the MAPE of day-of-the-week forecasting models that incorporate long memory behaviour by means of ARFIMA transformations or heterogeneous parameters. Panel A documents AR(1) alternatives whereas panel B classifies AR(9) alternatives. Diebold and Mariano (1995) statistics are included as an instrument for comparing models on a significance level, that is denoted *, ** and *** representing 10%, 5% and 1%, respectively. A positive DM statistic implies that the row model outperforms the column model, and vice versa. Both panels are further classified in 95% model confidence sets, and underlined metrics indicate presence in these sets. Metrics in **bold** indicate that, according to that metric, the model is present in the combined model confidence set comprising all panels.

Table 7: Forecast performances with long memory components

specifications are considered.

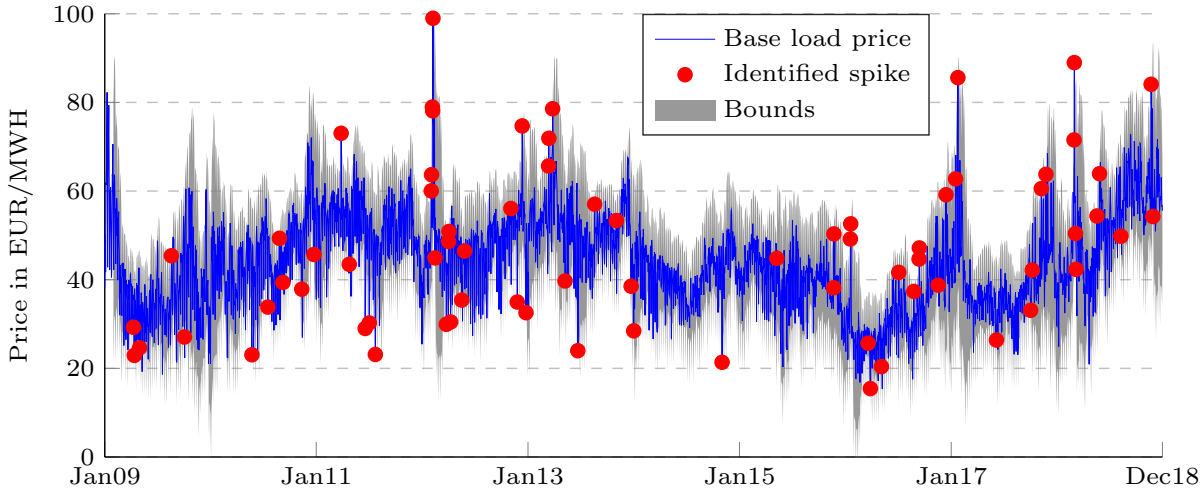
On the basis of the aggregated model confidence set, it is striking that AR(1) models equipped with heterogeneous parameters perform relatively equal to the AR(9) alternatives. In particular, the H(7,30)AR(1) model seems to perform similar to the H(30)AR(9) model, even though it contains 70% less parameters. Whereas ARFIMA alternatives come with substantial higher complexity, including heterogeneous parameters does not. Therefore, an important conclusion from this section is that the H(7,30)AR(1) day-of-the-week specification is a better alternative compared to the day-of-the-week AR(9), as it simply comprises less parameters. However, it is not unambiguous that this conclusion holds when spikes are incorporated or when technical optimization methods are applied. Consequently, the remainder of this study will focus on the H(7,30)AR(1) specification primarily but report quantitative results for the AR(9) model as well.

5.3 Spikes

Although the simulation analysis on the filter's identification ability was conclusive, the impact of applying the algorithm in forecasting models is yet to be determined. Figure 10 illustrates the 72 identified spikes by the limiting approach, together with the confidence interval of normal prices. It follows that the algorithm has identified a reasonably number of spikes, and that most identifications are in line with graphical expectations. However, the analysis also shows spikes

that are in some instances grouped. In such periods, treating the spikes might actually increase prediction errors as subsequent prices are predicted to be even lower. The extent to which pre-filtering improves forecasting performance is therefore subject to a trade-off: applying extreme treatment schemes may improve performance for periods where prices are relatively normal, but deteriorate in periods where extreme prices are present.

Figure 10: Spike identification process

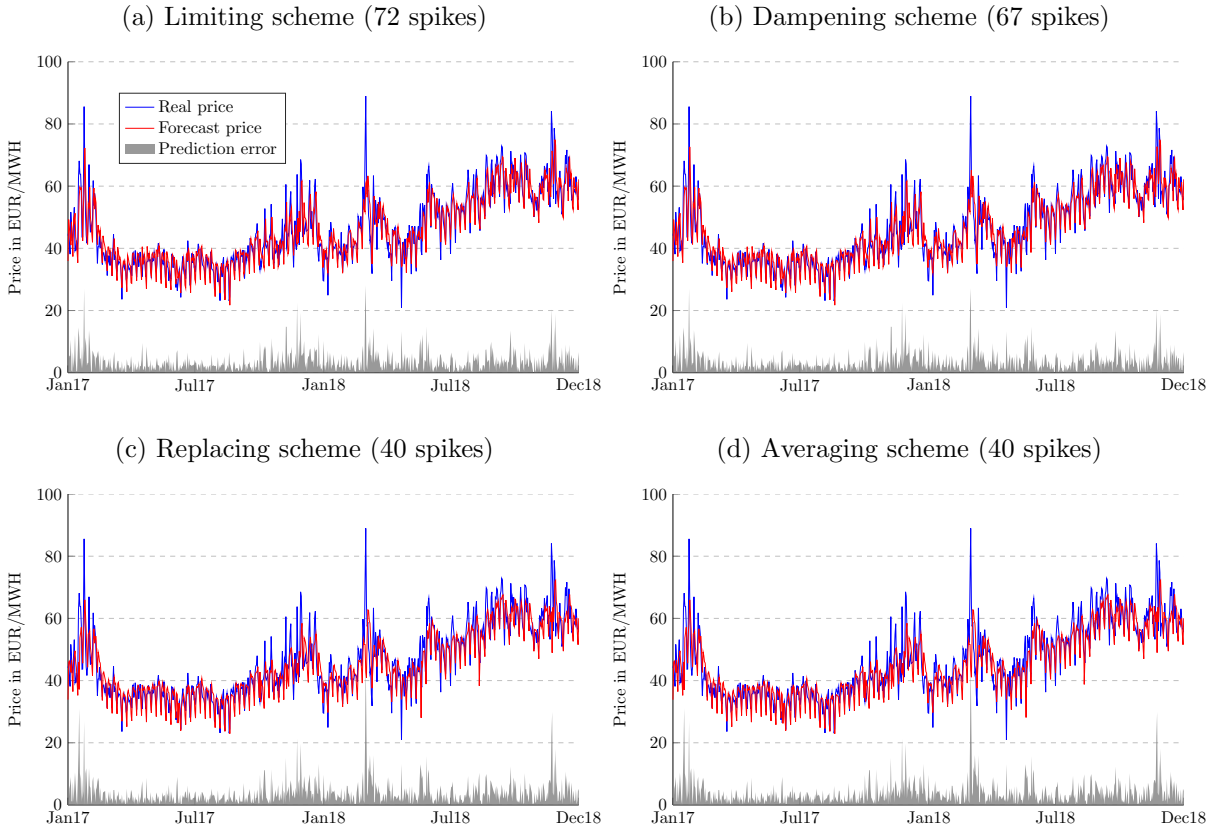


Notes: In this figure, the identified spikes (72) are presented as a red dot within a series of daily base load electricity prices ranging from January 2009 to December 2018. The identification is based on the iterative rolling window algorithm using the limiting scheme (Algorithm 1). The grey area represents the range to which normal prices are restricted.

Price spikes are treated using different methods, of which the predictive performances are presented in Figures 11a to 11d for the H(7,30)AR(1) day-of-the-week model. The replacing and averaging schemes evidently cause high prediction errors in periods with extreme prices. These two methods therefore treat spikes to aggressively and disturb the series too heavily. The limiting and dampening schemes seem to perform relatively equally. However, the performance of the four methods is less observable in periods without extremes. Therefore, a quantitative analysis is required.

Table 8 exhibits the predictive performances of spike-filtered forecasts, based on the two optimal models, H(7,30)AR(1) and AR(9), complemented with day-of-the-week effects. It is evident that none of the treatment techniques leads to significant improvements, consistent for both models. Moreover, the replacing and averaging methods yield significantly higher prediction errors than their unfiltered alternative, both being significant on 1% level. In line with the employed simulations, the replacing and averaging methods are unattractive for the purpose of forecasting. For each of the two models, and consistent for both the MAE and MAPE evaluation metrics, limiting or dampening the spikes is a better alternative, indicated by the MCS. However, both do not yield significant improvements relative to unfiltered models. On

Figure 11: H(7,30)AR(1) day-of-the-week forecasts with different spike treatment schemes



Notes: These four figures present the predictive accuracy of spike pre-processed electricity prices, based on four different techniques. All models are based on H(7,30)AR(1) specifications with day-of-the-week effects only.

the aggregated level, the MAE and MAPE both indicate that the H(7,30)AR(1) and AR(9) specifications perform equally in limiting or dampening spikes. It is striking, though, that in terms of the MSE, all treatment schemes are present in the combined MCS. A more thorough analysis showed that in this MCS procedure, most p-values are just above 10%, and consequently insignificant.

The corresponding hypothesis was defined as follows.

Hypothesis 4 *Dutch day-ahead electricity price predictions improve when spikes are filtered.*

In the Dutch day-ahead electricity market, price spikes cause impose challenges in forecasting models. But as observed, correcting for these spikes is difficult. Insufficient evidence is found for the claim that controlling for spikes improves forecasting performance. In fact, some techniques even yield higher prediction errors. Although the limiting and dampening schemes outperform the replacing and averaging schemes in most cases, no significant improvements are observed compared to unfiltered models. Using Figure 10, periods in which prices climb rapidly over subsequent days occur regularly. Pre-filtering the prices does probably not sufficiently correct for these situations, as the iterative framework implicitly assumes that every spike is incidental

Treatment	MAE	MAE vs unfiltered	MSE	MSE vs unfiltered	MAPE	MAPE vs unfiltered
Panel A: H(7,30)AR(1)						
Limiting	**3.479**	-0.15	**23.377**	-0.99	**7.59%**	0.14
Dampening	**3.472**	0.72	**23.246**	-0.41	**7.58%**	0.98
Replacing	3.779	-3.85***	**30.352**	-2.59***	8.09%	-3.60***
Averaging	3.778	-3.86***	**30.281**	-2.60***	8.09%	-3.61***
Panel B: AR(9)						
Limiting	**3.500**	0.25	**23.377**	-0.21	**7.60%**	0.60
Dampening	**3.492**	1.10	**23.262**	0.49	**7.59%**	1.36*
Replacing	3.977	-4.12***	**36.877**	-2.78***	8.44%	-4.15***
Averaging	3.975	-4.12***	**36.702**	-2.80***	8.44%	-4.16***

Notes: This table presents the mean average error (MAE), the mean squared error (MSE) and the mean absolute percentage error (MAPE) for spike pre-filtered forecasts. Diebold and Mariano (1995) are reported indicating whether a specific treatment scheme yields lower prediction errors as compared to the base model, indicated with *, ** and ***, indicating 10%, 5% and 1% significance, respectively. Furthermore, 95% model confidence sets are constructed for each panel, where an underlined statistics implies presence in a particular set. Metrics documented in **bold** imply presence in the model confidence set combining all panels.

Table 8: Predictive performances of spike pre-filtered spot prices

and not succeeded by other spikes.

5.4 Optimization

The first optimization approach is adding moving average terms, of which Panel A (Table 9) documents the results. It is evident that adding $q = 1, 2, 3$ moving average lags to H(7,30)AR(1) models does not yield improvements. Both the MAE, MSE and MAPE are very similar to that of the H(7,30)AR(1), while the models become more complex in the process. For the AR(9) model, adding two moving average terms does in fact reduce prediction errors, consistently for all three metrics on a 5% level. The MAPE is decreased by 0.05 percent points.

Even though the series of day-ahead prices is far from normally distributed, logarithmic transformations are ineligible. In fact, for both models, logarithmic predictions yield significantly higher prediction errors on a 1% level, increasing the MAPE by three to four percent points (Panel B). Panel C further highlights that combining moving average terms and logarithmic transformations also produces much higher prediction errors. Logarithmic transformations are therefore undesirable.

Despite the higher number of parameters, the ARMA(9,2) day-of-the-week model outperforms the AR(9) alternative. Consequently, two specifications are considered optimal at this point: the ARMA(9,2) and H(7,30)AR(1) day-of-the-week models, without spike filters. Panel D evaluates both models by means the model confidence set approach, and documents that both

models perform equally. But since the H(7,30)AR(1) model is more parsimonious—it contains more than 70% less parameters—the final conclusion is that electricity prices are most accurately predicted with an AR(1) model that is extended with both day-of-the-week dummies and two heterogeneous parameters, representing a weekly and monthly dependency.

Model	MAE	MAE vs basic	MSE	MSE vs basic	MAPE	MAPE vs basic
Panel A: Moving average terms (vs. non-MA)						
H(7,30)ARMA(1,1)	3.478	-0.02	23.124	1.08	7.59%	0.02
H(7,30)ARMA(1,2)	3.479	-0.16	23.135	0.93	7.59%	-0.06
H(7,30)ARMA(1,3)	3.483	-0.79	23.169	0.46	7.60%	-0.62
ARMA(9,1)	3.491	1.26	23.236	1.11	7.61%	1.04
ARMA(9,2)	3.476	2.23**	23.014	2.10**	7.58%	1.85**
Panel B: Logarithms (vs. non-logarithms)						
H(7,30)AR(1) in logs	4.979	-11.73***	40.041	-9.66***	11.69%	-12.56***
AR(9) in logs	4.466	-8.97***	34.725	-7.42***	10.15%	-9.97***
Panel C: Combined						
H(7,30)ARMA(1,1) in logs	5.147		42.391		11.94%	
H(7,30)ARMA(1,2) in logs	4.706		35.855		10.84%	
H(7,30)ARMA(1,3) in logs	4.705		35.707		10.84%	
ARMA(9,1) in logs	4.467		34.730		10.15%	
ARMA(9,2) in logs	4.442		34.443		10.09%	
Panel D: Evaluation of optimal models						
H(7,30)AR(1)	3.478		23.203		7.59%	
ARMA(9,2)	3.476		23.014		7.58%	

Notes: This table documents the predictive performance of two optimisation techniques: moving average terms (Panel A) and logarithmic transformations (Panel B), as well as combinations of both (Panel C). All models include day-of-the-week dummies. Diebold and Mariano (1995) statistics are reported indicating whether the specified model outperforms its basic alternative. Significance is expressed by *, ** and ***, indicating a 10%, 5% or 1% level, respectively. To measure inter-model performance, 95% model confidence sets are constructed for each panel, where an underlined statistic refers to presence in that particular set based on that statistic. Panel D documents the performances of the optimal models as selected by the previous panels in combination with previous results.

Table 9: Predictive performances resulting from two optimisation techniques

6 Relation analysis

This chapter covers the relationship analysis between day-ahead electricity price predictability and activity on the intra-day imbalance market. Both the imbalance price volatility and the aggregated daily volume are regressed on three distinct predictability measures. Based on the theory regarding the purpose of the imbalance market, it is expected that price predictability on the day-ahead market is negatively related to both metrics. When prices are harder to predict for day t, uncertainty arises and the imbalance market is likely affected by a larger trading volume and price volatility.

6.1 Price volatility

Firstly, the imbalance price volatility, measured by the standard deviation, is analyzed from three angles. Table 10 documents the results of the regression models. Models I and II depict the relationship between the price volatility and the 24-hour inter-day correlation, $\rho(t)$, as a measure of day-ahead predictability. It is evident that in both models, the constant coefficient is strong and highly significant, constituting approximately €51.28 and €60.93 in average price deviation, respectively, for every 15-minute window. In the stand-alone model, $\rho(t)$ is only significant at a 10% level. However, when time fixed effects are included, its effect becomes stronger and significant on a 5% level. Consequently, from the perspective of predictability measure $\rho(t)$, evidence is found for a negative relationship with price volatility on the imbalance market. More specifically, the coefficient of -5.790 implies that every decrease of 1 in the correlation coefficient of the 24-hour day-ahead prices between day $t-1$ and day t, coincides with an average increase of €5.79 per megawatt hour in 15-minute price volatility on the imbalance market on day t. It should be noted, though, that the R^2 measure is very small for $\rho(t)$ only—the 24-hour correlation does not explain much variation in price volatility. Further, including control variables leads to strong increases in the (adjusted) R^2, which indicates that time fixed effects are present.

The second predictability measure, $\gamma(t)$, is a broader metric of correlation, and is calculated by the first-lag serial correlation of base load prices within a window $t-19, \cdots, t$ with length 20. Similarly to the previous models, the constant coefficient in models III and IV is high and significant on a 1% level. However, $\gamma(t)$ is insignificant and even positive in both the stand-alone model and the model supplemented with control variables. Combined with very low (adjusted) R^2 metrics, insufficient evidence is found for the existence of a relationship between measure $\gamma(t)$ and 15-minute price volatility on the imbalance market.

Lastly, model-implied predictability is employed as an instrument for measuring predictability on the day-ahead market, based on the optimal H(7,30)AR(1) model with day-of-the-week effects. Models V and VI present a marginal positive and insignificant coefficient, as well as low R^2 measures. Therefore, no evidence is found for a negative relationship between the model implied predictability and price volatility on the imbalance market.

Volatility	Model					
	I	II	III	IV	V	VI
α	51.282***	60.933***	47.990***	55.886***	44.732***	59.151***
	(1.425)	(2.438)	(0.915)	(2.254)	(1.409)	(15.254)
$\rho(t)$	-3.602*	-5.790**				
	(1.872)	(2.510)				
$\gamma(t)$			1.830	2.463		
			(2.435)	(2.444)		
$\ell(t)$					1.028	0.515
					(0.346)	(0.344)
Time fixed effects	×	✓	×	✓	×	✓
R^2	0.0009	0.1018	0.0002	0.1002	0.0193	0.1281
Adjusted R^2	0.0007	0.0971	-0.0001	0.0955	0.0179	0.1047
Observations	3652	3652	3633	3633	730	730

Notes: This table documents the results of six regression models. For every predictability measure, $\rho(t)$, $\gamma(t)$ and $\ell(t)$, defined in Equation 25, 27 and 28, respectively, two models are estimated. The first model includes only the measure and a constant, whereas the second also contains time fixed effects (month-of-the-year dummies, day-of-the-week dummies and day-specific parameter t) as control instruments. White (1980) standard errors are mentioned between parentheses below the coefficient estimates. Significance on a 10%, 5% and 1% level is denoted *, ** and ***, respectively.

Table 10: Regressions with daily imbalance price volatility as dependent variable

The corresponding hypothesis was formulated as follows:

Hypothesis 5 *The predictability of Dutch day-ahead electricity prices relates negatively to price volatility on the imbalance market.*

Comparing all three predictability metrics, the hourly price inter-day correlation was argued most reliable, as it incorporates real established prices instead of statistical aggregated base load prices. Furthermore, it is day-specific and therefore very precise, in contrast to the rolling window autocorrelation approach. Predictability implied by prediction errors was assumed as the least reliable measure, as it is subject to model risk.

In conclusion, a negative relationship was found based on the hourly inter-day correlation. The other two metrics are clearly undesirable in measuring predictability. Therefore, sufficient evidence is found to support the claim of this hypothesis. On average, a lower (higher) day-ahead price predictability, as based on the correlation with the hourly prices of the previous day, coincides with greater (smaller) imbalance price volatility on that day.

6.2 Market volume

Next to prices on the imbalance market, quantities are analyzed. When day-ahead price predictability is lower for a specific day t, it is likely that volumes traded on the day-ahead market are also uncertain. In that case, the imbalance market is, indirectly, expected to be subject to higher trade due to mismatches on the day-ahead market.

Inter-day correlation of hourly day-ahead prices between $t-1$ and t, expressed $\rho(t)$, is analyzed first. Models I and II in Table 11 document significant constants. On average, with (without) time fixed effects and assuming $\rho(t) = 0$, approximately 11.859 (9.370) gigawatt hours are traded on a daily basis. Whereas a negative significant coefficient -2.162 is reported for $\rho(t)$ in the stand-alone model, controlling for time fixed effects reduces its influence to an insignificant coefficient -0.399. Since the (adjusted) R^2 increases substantially when time fixed effects are added, model II is clearly more relevant than model I. Therefore, insufficient evidence is found for the existence of a consistent relationship between $\rho(t)$ and the daily imbalance market volume.

Secondly, first-lag serial correlation on a rolling window is used as a proxy for day-ahead predictability. Constants α are significant and similar to models I and II. Surprisingly, $\gamma(t)$ is found significantly positive in the stand-alone specification. Correcting for time fixed effects increases its coefficient even more. Measuring day-ahead price predictability by means of $\gamma(t)$ therefore seems positively related to aggregated daily volumes on the imbalance market. Controlled for time fixed effects, an increase of 1 in the rolling window autocorrelation on average coincides with an increase of approximately 1.625 gigawatt hours in total daily volume on the imbalance market. Again, the R^2 is very small, below 1%, for the $\gamma(t)$ coefficient individually. The observed positive relationship is rather puzzling. A high $\gamma(t)$ implies that, on average, day-ahead prices in the window $t-18, \cdots, t$ are strongly and positively dependent on their preceding price. However, this does not necessarily reflect the predictability for day t specifically. This approach can consequently be interpreted as indirect and possibly inaccurate, substantiated by a very low correlation with $\rho(t)$ of approximately 0.05. The observed positive relationship is illustrated by means of scatter plots (Appendix D). Although some outliers are present, removing these does not substantially reduce the positive relationship, for both the standalone case and the case where time fixed effects are filtered out. At least, the relationship is not negative, but the positive relationship cannot be elucidated either.

Lastly, day-ahead price predictability is represented by the mean absolute error of the H(7,30)AR(1) model, of which the results are presented in models V and VI. Both the standalone model and the model including time fixed effects document no significant relationship. Emphasized by a standalone R^2 of less than 1%, using the prediction errors as a measure for

predictability is clearly undesirable. Since market participants likely employ very different and more complex models, not necessarily driven by time series, model risk clearly materializes.

Volume	Model					
	I	II	III	IV	V	VI
α	11.859***	9.370***	9.747***	8.312***	13.618***	5.408***
	(0.191)	(0.282)	(0.115)	(0.248)	(0.170)	(1.732)
$\rho(t)$	-2.162***	-0.399				
	(0.247)	(0.274)				
$\gamma(t)$			1.546***	1.625***		
			(0.321)	(0.293)		
$\ell(t)$					0.041	-0.013
					(0.037)	(0.038)
Time fixed effects	×	✓	×	✓	×	✓
R^2	0.0194	0.2396	0.0064	0.2463	0.0023	0.1220
Adjusted R^2	0.0192	0.2356	0.0061	0.2424	0.0009	0.0985
Observations	3652	3652	3633	3633	730	730

Notes: In this table, coefficient estimations are presented with the aggregated daily volume, $V_t^{\asymp}$, as the regressand, and different predictability measures, $\rho(t)$, $\gamma(t)$ and $\ell(t)$, as regressors. Time fixed effects consisting of month-of-the-year dummies, day-of-the-week dummies and a daily factor $t = 1, \cdots, T$, are added as control variables. Robust White (1980) standard errors are documented between parentheses below coefficient estimates. Significant coefficients are appended with *, ** or ***, corresponding to significance on a 10%, 5% or 1% level, respectively.

Table 11: Regression models with aggregated daily imbalance volume as dependent variable

The expected relationship between day-ahead price predictability and daily volumes on the imbalance market was formulated by the following hypothesis:

Hypothesis 6 *Daily volumes on the Dutch imbalance market negatively relate to price predictability on the day-ahead market.*

As elaborated on in the previous section, inter-day hourly price correlation is considered the optimal metric to proxy for predictability on the day-ahead market. Through a significant negative coefficient, evidence is found for a negative relationship to exist. However, its effect is heavily reduced to an insignificant coefficient when control variables are included. Moreover, using the rolling window autocorrelation metric exhibits a consistent positive relationship with predictability, which is puzzling at least. Consequently, insufficient evidence is found for this hypothesis and for a significant negative relationship to exist. The most appealing cause for the absence of a strong negative relationship is that market players rapidly adjust their capacity after strong demand or supply shocks. As a result, PTU-specific shocks would be compensated by inverse reactions and aggregating these on a daily basis would cancel out these shocks. Using the daily averaged PTU imbalance might yield a more reliable estimate. However, this would presumably yield very similar results to the price volatility.

7 Conclusion

This study focuses on Dutch electricity prices. Most importantly, it attempts to fill the academic gap in predicting day-ahead prices. Next to that, the imbalance market is analyzed by examining its relationship with day-ahead price predictability. The research question was formulated as follows:

Research Question *To what extent are Dutch electricity day-ahead prices predictable and how does predictability relate to intra-day imbalance trading?*

Dutch day-ahead prices ranging from 2009 to 2018 are examined in a predictive framework, based on AR(F)IMA type specifications supplemented with various extensions. Strong autocorrelation effects are observed, and the autocovariance structure is therefore leading. Extending different types of models with day-of-the-week dummies enhances the predictive performance. However, incorporating month-of-the-year effects does not. Since autocorrelation is present up to a high number of lags, long-memory properties are modelled using ARFIMA and HAR components. Another key property of electricity prices is the occurrence of spikes. This study proposes a new pre-filter algorithm that aims towards improving forecasting models. Although simulations suggest that the algorithm is accurate using multiple techniques, significant forecast improvements are not achieved. Based on performance as well as on parsimony, the H(7,30)AR(1) model with day-of-the-week effects is found optimal for predictive processes, with a MAE, MSE and MAPE of 3.478, 23.203 and 7.59%, respectively. It follows that, on average, day-ahead prices are predictable for approximately 92%. These results are relatively in line with both Cuaresma et al. (2004) and Kristiansen (2012), analyzing the Leipzig Power Exchange and Nordic power market, respectively. Whereas Raviv et al. (2015) observe that HAR models do not provide any additional power relative to the other models, the present study finds that HAR specifications represent better and less complex alternatives to high-order AR models. Overall, it is evident that the dynamics on the Dutch market are relatively in line with other markets, and that time series models are appropriate in a predictive context. This study contributes to the academic foundation of the Dutch market, and helps in building predictive pricing models.

Day-ahead price predictability is expected to influence the imbalance market, on which real-time surpluses and shortages are traded. Whenever day-ahead prices are difficult to predict and supply and/or demand suffer from uncertainty, the imbalance market is expected to be needed for compensation. Using the inter-day correlation of hourly day-ahead prices as a proxy for

predictability, a negative relationship is observed with daily price volatility on the imbalance market. Other things equal, when day-ahead prices are highly predictable (unpredictable), imbalance prices are relatively stable (volatile). The aggregated daily volume on the imbalance market is not found negatively related to day-ahead predictability. In terms of volume, the imbalance market is therefore found relatively efficient, as PTU-specific supply and demand shocks do not have a substantial impact on the daily aggregated volume. In conclusion, this study finds that the day-ahead market and imbalance market are related and that the lower the day-ahead price predictability, the more price volatility emerges on the imbalance market. The economic implication is that trading on the imbalance market can be employed as a profitable strategy when day-ahead price predictability is low. The relationship between these markets also emphasizes the relevance of having accurate prediction models.

The objective of this study is to build an academic foundation for Dutch electricity prices, using a relevant and recent data set. However, since the analysis is mainly based on prediction results, price dynamics are not studied in detail in terms of their relevance in-sample. This represents a clear limitation. In order to disentangle the presence of the day-of-the-week effect or month-of-the-year effect, it would be helpful to study the dummy coefficients individually. Furthermore, it is crucial to note that the models that are employed are relatively simple. Most importantly, they comprise only the autocovariance structure, time fixed effects and long-memory components. In order to improve the forecasts, future studies could add additional exogenous variables that contain valuable price information (ARMAX), such as the predicted load or weather conditions (Kristiansen, 2012). The models could also be improved by explicitly allowing a time-varying volatility structure using GARCH effects (Garcia et al., 2005), or decomposing the series by a wavelet transform methodology to attain sub-series that can be modelled more effectively (Tan et al., 2010). Especially when analyzing large samples, model combination techniques could be employed to allow for time-varying dynamics (Raviv et al., 2015). Finally, as Cuaresma et al. (2004) and Raviv et al. (2015) argue, predicting hourly price series individually is likely to improve the results relative to predicting base load prices, mainly due to the fact that hourly prices are actually determined by the market system and exhibit a specific mean and volatility structure (Huisman et al., 2007).

References

Akaike, H. (1974). Stochastic theory of minimal realization. *IEEE Transactions on Automatic Control*, *19*(6), 667–674.

Amjady, N. (2006). Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Transactions on Power Systems*, *21*(2), 887–896.

Andersen, T. G., Bollerslev, T., & Dobrev, D. (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics*, *138*(1), 125–180.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, *73*(1), 5–59.

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.

Bordignon, S., Bunn, D. W., Lisi, F., & Nan, F. (2013). Combining day-ahead forecasts for British electricity prices. *Energy Economics*, *35*, 88–103.

Borenstein, S. (2002). The Trouble With Electricity Markets: Understanding California's Restructuring Disaster. *Journal of Economic Perspectives*, *16*(1), 191–211.

Bowden, N., & Payne, J. E. (2008). Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models. *Energy Economics*, *30*(6), 3186–3197.

Brooks, C. (2014). *Introductory econometrics for finance* (3rd ed.). Cambridge University Press.

Clewlow, L., & Strickland, C. (2000). *Energy derivatives: Pricing and risk management*. Lacima Publications.

Clò, S., Cataldi, A., & Zoppoli, P. (2015). The merit-order effect in the Italian power market: The impact of solar and wind generation on national wholesale electricity prices. *Energy Policy*, *77*, 79–88.

Conejo, A. J., Plazas, M. A., Espinola, R., & Molina, A. B. (2005). Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*, *20*(2), 1035–1042.

Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, *18*(3), 1014–1020.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Cuaresma, J. C., Hlouskova, J., Kossmeier, S., & Obersteiner, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, *77*(1), 87–106.

de Menezes, L. M., Houllier, M. A., & Tamvakis, M. (2016). Time-varying convergence in European electricity spot markets and their association with carbon and fuel prices. *Energy Policy*, *88*, 613–627.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427–431.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

EPEX SPOT SE. (2018). *Trading on EPEX SPOT 2018-2019.* Retrieved from https://www.epexspot.com/document/39262/2018-05-28_TradingBrochure_web.pdf

European Union. (2009). Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC. *Official Journal of the European Union*, *5*, 16–62.

Farahmand, H., & Doorman, G. (2012). Balancing market integration in the Northern European continent. *Applied Energy*, *96*, 316–326. (Smart Grids)

Garcia, R. C., Contreras, J., van Akkeren, M., & Garcia, J. B. C. (2005). A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, *20*(2), 867–874.

Gianfreda, A., & Grossi, L. (2012). Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics*, *34*(6), 2228–2239.

Gough, R., Dickerson, C., Rowley, P., & Walsh, C. (2017). Vehicle-to-grid feasibility: A techno-economic analysis of ev-based energy storage. *Applied Energy*, *192*, 12–23.

Haldrup, N., & Ørregaard Nielsen, M. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, *135*(1), 349–376.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, *68*(1), 165–176.

Huisman, R. (2008). The influence of temperature on spike probability in day-ahead power prices. *Energy Economics*, *30*(5), 2697–2704.

Huisman, R., Huurman, C., & Mahieu, R. (2007). Hourly electricity prices in day-ahead markets. *Energy Economics*, *29*(2), 240–248.

Huisman, R., & Kiliç, M. (2013). A history of European electricity day-ahead prices. *Applied Economics*, *45*(18), 2683–2693.

Huisman, R., & Mahieu, R. (2003). Regime jumps in electricity prices. *Energy Economics*, *25*(5), 425–434.

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

Keles, D., Genoese, M., Möst, D., & Fichtner, W. (2012). Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. *Energy Economics*, *34*(4), 1012–1032.

Knittel, C. R., & Roberts, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, *27*(5), 791–817.

Kristiansen, T. (2012). Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy*, *49*, 328–332.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1), 159–178.

Lago, J., Ridder, F. D., Vrancx, P., & Schutter, B. D. (2018). Forecasting day-ahead electricity prices in Europe: The importance of considering market integration. *Applied Energy*, *211*, 890–903.

Lee, D., & Schmidt, P. (1996). On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics*, *73*(1), 285–302.

Lee, S. S., & Mykland, P. A. (2007, 12). Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *The Review of Financial Studies*, *21*(6), 2535–2563.

Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, *37*, 152–166.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*(1), 125–144.

Mount, T. D., Ning, Y., & Cai, X. (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, *28*(1), 62–80.

Mulder, M., & Schoonbeek, L. (2013). Decomposing changes in competition in the Dutch electricity market through the residual supply index. *Energy Economics*, *39*, 100–107.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Nogales, F. J., Contreras, J., Conejo, A. J., & Espinola, R. (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, *17*(2), 342–348.

Nowotarski, J., & Weron, R. (2016). On the importance of the long-term seasonal component in day-ahead electricity price forecasting. *Energy Economics*, *57*, 228–235.

Phillips, P. C. B., & Perron, P. (1988, 06). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335–346.

Purvins, A., Zubaryeva, A., Llorente, M., Tzimas, E., & Mercier, A. (2011). Challenges and options for a large wind power uptake by the European electricity system. *Applied Energy*, *88*(5), 1461–1469.

Raviv, E., Bouwman, K. E., & van Dijk, D. (2015). Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics*, *50*, 227–239.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Shahidehpour, M., Yamin, H., & Li, Z. (2002). *Market operations in electric power systems: Forecasting, scheduling, and risk management*. Wiley.

Tan, Z., Zhang, J., Wang, J., & Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy*, *87*(11), 3606–3610.

TenneT Holding B.V. (2016). *The Imbalance Pricing System* . Retrieved from https://www.tennet.eu/fileadmin/user_upload/Company/Publications/Technical_Publications/Dutch/imbalanceprice_3.6_clean_.doc.pdf

TenneT Holding B.V. (2018). *TenneT Market Review 2017 – Electricity market insights.* Retrieved from https://www.tennet.eu/fileadmin/user_upload/Company/Publications/Technical_Publications/Dutch/2017_TenneT_Market_Review.pdf

Weron, R. (2006). *Modeling and forecasting electricity loads and prices: A statistical approach*. John Wiley & Sons.

Weron, R. (2008). Market price of risk implied by Asian-style electricity options and futures. *Energy Economics*, *30*(3), 1098–1115.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838.

Zachmann, G. (2008). Electricity wholesale market prices in Europe: Convergence? *Energy Economics*, *30*(4), 1659–1671.

Appendices

A Prices and returns

Model	Prices			Returns		
	MAE	MSE	MAPE	MAE	MSE	MAPE
AR(1)	4.287	35.424	9.53%	4.498	38.246	9.96%
AR(2)	4.283	35.254	9.52%	4.411	34.915	9.80%
AR(3)	4.170	32.782	9.25%	4.360	33.997	9.69%
AR(4)	4.149	32.286	9.21%	4.354	33.761	9.71%
AR(5)	4.141	31.996	9.20%	4.197	31.654	9.34%
AR(6)	3.994	29.952	8.83%	4.168	31.620	9.19%
AR(7)	3.896	29.111	8.53%	4.027	30.887	8.85%
AR(8)	3.825	28.619	8.35%	4.042	31.217	8.87%
AR(9)	3.841	28.842	8.38%	4.043	31.209	8.87%
AR(10)	3.843	28.839	8.38%	4.047	31.234	8.88%

Notes: Out-of-sample prediction metrics are reported for both prices and returns, to determine if modelling prices is justified as a non-$I(0)$ series.

Table 12: Out-of-sample performance criteria for both prices and returns

B Information criteria

Model	Full sample		In-sample	
	AIC	BIC	AIC	BIC
Panel A: AR(p)				
AR(1)	23299.277	23317.886	18616.519	18634.459
AR(2)	23295.289	23320.101	18615.617	18639.537
AR(3)	23096.321	23127.336	18472.324	18502.224
AR(4)	23045.278	23082.496	18433.345	18469.226
AR(5)	22945.015	22988.436	18339.737	18381.597
AR(6)	22446.477	22496.101	17887.803	17935.644
AR(7)	22075.595	22131.422	17531.124	17584.945
AR(8)	21862.380	21924.410	17318.195	17377.995
AR(9)	21854.285	21922.518	17304.999	17370.779
AR(10)	21856.049	21930.486	17306.879	17378.640
Panel B: ARMA(p,q)				
ARMA(1,1)	23286.616	23311.428	18611.828	18635.748
ARMA(1,2)	22713.924	22744.939	18153.383	18183.283
ARMA(1,3)	22695.512	22732.730	18138.670	18174.550
ARMA(9,1)	21856.097	21930.533	17306.890	17378.650
ARMA(9,2)	21842.172	21922.811	17294.278	17372.018
ARMA(9,3)	21843.644	21930.487	17296.095	17379.815
Panel C: ARFIMA(p,d,q)				
ARFIMA(1,d,0)	22884.660	22909.470	18287.210	18311.130
ARFIMA(1,d,1)	22797.260	22828.280	18217.260	18247.160
ARFIMA(1,d,2)	22727.490	22764.710	18158.230	18194.110
ARFIMA(1,d,3)	22778.630	22822.060	18206.690	18248.550
ARFIMA(9,d,0)	21828.180	21902.620	17282.690	17354.450
ARFIMA(9,d,1)	21821.760	21902.400	17275.370	17353.110
ARFIMA(9,d,2)	21809.640	21896.480	17264.120	17347.840
ARFIMA(9,d,3)	21811.520	21904.570	17265.740	17355.440

Notes: This table documents Akaike (1974) and Schwarz (1978) (Bayesian) information criteria, abbreviated AIC and BIC, respectively, for three model classes. A lower metric indicates a better model that potentially yields higher predictive accuracy. In the AR(p) family, $p = 9$ performs best and panel B and C are therefore based on nine autoregressive parameters.

Table 13: Model comparisons for different classes based on information criteria

C Spike simulation DGP

Pseudo asset prices are simulated from 2009 up to 2018 ($T = 3,652$), based on the simulation setup of Janczura et al. (2013). Essentially, both the long-term component and time fixed effects are extracted and retained. The resulting stochastic residuals are simulated to generate a fictive process. The main advantage of this approach is that its structure is based on real prices, thereby providing a realistic price process.

First, the long-term seasonal component is described by a sinusoidal combined with an exponentially weighted moving average (EWMA) component, where the latter is computed by

$$\text{EWMA}_t^\lambda = (1-\lambda)P_t + \lambda \text{EWMA}_{t-1}^\lambda, \tag{31}$$

where EWMA_1^λ is set to the average price over the whole sample. Then, the long-term trajectory T_t is formally defined by

$$T_t = \alpha_1 \sin\left[2\pi\left(\frac{t}{365} + \alpha_2\right)\right] + \alpha_3 + \alpha_4 \text{EWMA}_t^\lambda, \tag{32}$$

where decay factor $\lambda = 0.975$ and coefficients α_i are estimated using a nonlinear regression.

Subsequently, the day-of-the-week component s_t is extracted by regressing $P_t - \hat{T}_t = \alpha + \sum_{i=2}^{7} \beta_i I_{t,i} + \varepsilon_t$ and computing

$$\hat{s}_t = \hat{\alpha} + \sum_{i=2}^{7} \hat{\beta}_i I_{t,i}, \tag{33}$$

where $I_{t,i}$ denotes a dummy variable indicating whether the day of the week of observation t is i. The stochastic price component is then derived as

$$\hat{X}_t = P_t - \hat{T}_t - \hat{s}_t. \tag{34}$$

Based on Janczura et al. (2013), extreme observations X_t exceeding a two-sided 5% confidence interval are replaced by the mean of the interval. Further, the stochastic component X_t is modelled by a mean reverting process that is described by

$$X_t = \alpha + (1-\beta)X_{t-1} + \sigma\varepsilon_t, \tag{35}$$

where σ denotes the standard deviation. Coefficients α and β are calibrated as $\hat{\alpha} \approx 0.001$ and $\hat{\beta} \approx 0.316$ using real prices. Based on these coefficients, the stochastic component is generated

by

$$dX_t = (\alpha - \beta X_t)dt + 0.8\sigma dW_t, \tag{36}$$

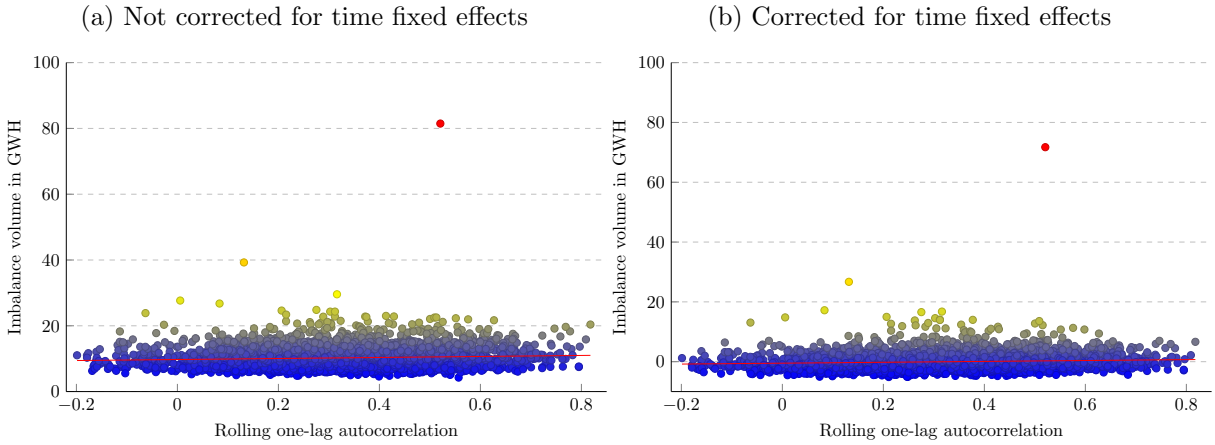
where $dt = 1$ (daily frequency) and $W_t \sim N(0, 1)$ denotes a Brownian motion process. The volatility component is reduced by 20% to obtain a more stable price series, that is more in line with the Dutch market and allows for more effective testing of the pre-filter algorithm.

Based on the stochastic process described by Equation 36, a price series is generated by adding the estimated seasonal and time fixed components. That is

$$P_t = \hat{T}_t + \hat{s}_t + X_t. \tag{37}$$

D Relation analysis

Figure 12: Correlation analysis between $\gamma(t)$ and $V_t^{\asymp}$



Notes: Figures (a) and (b) illustrate the correlation between day-ahead price predictability based on the one-lag autocorrelation within a rolling window of twenty days (x-axis) and the aggregated daily imbalance volume (y-axis). It is evident that the correlation is positive, also when time fixed effects are filtered out.