# Master's Thesis : Improvement of decision making for trading in wholesale electricity market

**Auteur :** Moitroux, Olivier
**Promoteur(s) :** Ernst, Damien
**Faculté :** Faculté des Sciences appliquées
**Diplôme :** Master en ingénieur civil en informatique, à finalité spécialisée en "intelligent systems"
**Année académique :** 2019-2020
**URI/URL :** /; http://hdl.handle.net/2268.2/9063

LIÈGE université
Sciences Appliquées

Master in applied sciences
Academic Year 2019-2020

LIÈGE université

# Improvement of decision making for trading in wholesale electricity market

A probabilistic forecasting approach

Master thesis carried out with a view to obtaining the master's degree in *Civil Engineering in Computer Science, specialising in intelligent systems*

**Author**
Olivier MOITROUX

**Supervisor**
Damien ERNST

Liège

June 2020

**Abstract**

It is common practice for risk-averse industrial companies to reduce their exposure to the volatile prices of the spot market by securing a base load supply on the year-ahead electricity market. While many research efforts have been put in designing strategies to interact with several markets and assets, the case of small industrial consumers bound to a block-size constrained click-by-click contract for the year-ahead market is overlooked in literature. This Master thesis seeks to explore this gap and aims at improving the purchase decision making process of such electricity consumers. Multivariate probabilistic forecasting is investigated as a mean to complement the trader's expertise. Compelling results are that year-ahead electricity prices expose random-like patterns which make future price inference extremely difficult. A comparison study of several time series model suggests that training global deep learning models on related time series noticeably improves the forecast accuracy but that simpler models produce better calibrated prediction intervals.

# Contents

# List of Figures

# List of Tables

# Acronyms

**ADF** Augmented Dickey-Fuller. 33, 34

**API** Application Programming Interface. 38, 43

**AR** Autoregressive. 36

**Belpex** Belgian Power Exchange. 14

**CAL** Calendar. 31

**DAM** Day Ahead Market. 8–12, 33, 56, 59

**DSO** Distribution System Operators. 4, 6

**EEX** European Energy Exchange. 6, 7, 57

**ENTSO-E** European Network of Transmission System Operators. 4, 9, 38

**EPF** Electricity Price Forecasting. 19, 20

**EWMA** Exponentially Weighted Moving Average. 29, 59

**FFNN** Feed Forward Neural Network. 23, 47

**GefCom2014** Global Energy Forecasting Competition 2014. 20, 26, 43

**GenCo** Generation Companies. 4, 9, 12

**HVDC** High Voltage Direct Current. 4

**IQR** Interquartile Range. 28

**LOESS** Locally Estimated Scatterplot Smoothing. 29, 58

**LSTM** Long-Short-Term-Memory. 21

**MCP** Market Clearing Price. 9

**MO** Market Operator. 6

**MOE** *Market Opportunity Estimator*. 2, 17

**MOI** *Market Opportunity Indicator*. 17

**MSIS** Mean Scaled Interval Score. 45

**OTC** Over The Counter. 6, 8, 13

**PI** Prediction Interval. 20, 45–48

**PICP** Prediction Interval Coverage Probability. 46

**QR** Quantile Regression. 26

**RES** Renewable Energy Sources. 5

**RNN** Recurrent Neural Network. 21

**Seq2Seq** Sequence-to-Sequence. 21

**SMP** System Marginal Price. 9

**SSM** State Space Model. 25

**SWDE** Société Wallonne Des Eaux. 2

**TLCC** *Time Lagged Cross Correlation.* 37

**TSO** Transmission System Operators. 4, 6, 8

**VaR** *Value-at-Risk.* 19, 58

**YAM** Year Ahead Market. 10–12, 16, 36, 38, 41, 62

# 1 Introduction

Electricity, due to its intrinsic nature, is one the most arduous commodity to exchange. Yet, the efforts put into the conception of complex pieces of engineering have led to the reliable electrical infrastructure that we know now and which we even tend to take for granted. Beside a physical infrastructure supporting its production and delivery, a set of markets has naturally evolved to trade the commodity. Among them, the wholesale electricity market gathers a set of market participants interested in purchasing and selling electricity ahead-of-delivery in large volumes. The year-ahead market is more specifically the one of interest for this Master's thesis.

In this context, the goal of this Master's thesis is to design an algorithmic solution to reduce the electricity acquisition cost for a small industrial company concerned with the improvement of its purchase decision making process in the year-ahead electricity market. An example of such company is the Société Wallonne Des Eaux (SWDE)[14]: a Belgian company which supplied water and sanitation services for a population of 2,459,369 people and which delivered 166,707,379 cubic meters of water in 2018[15]. While residential and small industrial consumers generally have a fixed-tariff contract with a given retailer, the SWDE is interested in trading directly in the wholesale electricy market to reduce its cost of energy acquisition. In practice, small industrial consumers cannot interact directly with the wholesale market but rather settle contracts by the intermediate of a supplier like Engie[16] for example.

The SWDE is willing to stabilize the price of water for its customers. For several years now, the company has strived to maintain the access to water at a fixed price despite the natural ongoing inflation. *Forward trading* is a technique which aims at alleviating the company's risks by reducing its exposure to the more volatile prices of short-delivery-horizon (wholesale) markets. Risk-averse companies prefer therefore this strategy, by covering, for example, 80% of their electricity demand on long-term-delivery markets. The SWDE is no exception and leverages derivative contracts on the *year-ahead market* in order to hedge against risk. The current situation is such that many managers in small industrial companies often rely on a loose human expertise to determine a sketch of their purchase timing.

This Master thesis thus aims at improving the quality of the decision making process for such generic company's profile. To this end, the research question to be addressed is the following:

"*At any given time of the quotation period, is it better to buy now or later?*"

Most of the corporate energy managers are still left with relative autonomy to sketch their purchase timing. This reason, coupled with the potentially very large traded amounts, makes resolving the problem with a fully end-to-end automatic algorithmic trading agent either too rigid or miss-aligned with the business reality. Instead, we aim at developing an approach based on an informative indicator to support the company's managers in their decision process. Furthermore, it is expected that the solution should be relatively *sustainable* and, likewise, easy to maintain in the future.

This thesis is organized as follows. First, Chapter 2 presents an introductory overview of the underlying mechanics of the electrical system with a focus on its financial aspects. We frame the problem in a formal way in Chapter 3. A literature review is conducted in Chapter 4 to introduce a set of candidate algorithms suitable for resolving the problem in a flexible way. Chapter 5 presents an in depth analysis of the data at hand so as to identify the possible challenges. In the second to last Chapter 6, we detail an evaluation methodology before providing a discussion on the results obtained. We finally conclude in Chapter 7.

# 2 Overview of the electrical system in a European and Belgian context

The European electricity system is a fascinating, yet very complex, ecosystem which gathers many actors around a physical infrastructure. In this way, it can be analyzed with two lenses. First, it can be seen as a network-like infrastructure, the *physical grid* which consists in generators, electricity transport systems as well as a distribution system. Second, it can be seen as a set of *organized markets* which allows economic agents to interact with each others in a controlled and safe manner. Complexity arises from the observation that electricity is a commodity which has the property that generation has to match consumption (plus grid losses) in real time [17]. Furthermore, it is not economically viable to store electricity at a large scale, which motivates a careful design of consecutive electricity markets.

The electricity system has recently undergone structural changes while also being more and more disturbed by new climate-related regulations which have impacted and complicated its dynamics. In the following sections, we aim at providing the reader with the relevant high-level and introductory concepts about electricity markets and their core-dynamics.

## 1 Electrical grid infrastructure

The grid consists in a network connecting electricity generators and consumers via transmission and distribution networks. Two very important properties characterize this network. First, supply and demand must be balanced at all time. A failure to meet this property would negatively impact the grid frequency and result in a power outage, also referred as a blackout[18]. Second, since electricity follows the path of lowest resistance, the flow of electricity within the grid cannot be easily controlled. As a result, consumers receive electricity from mixed sources. Figure (2.1) depicts a schematic overview of the physical actors involved in the system, whose roles will be briefly discussed. A more detailed schematic overview of an electrical grid with voltage and capacity figures is available in the Appendix 1 (Figure (8.1)).

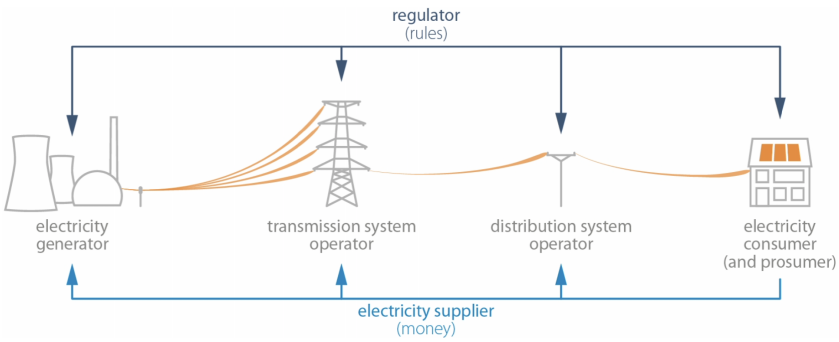



Figure 2.1 – Schematic overview of the electrical system [1]

## 1.1 Generators

Generators, or Generation Companies (GenCo), are the entities that own a single or a portfolio of plants which produce electricity. We distinguish two types of generators based on their capacity [1]. *Firm capacity* generators form the first category and are characterized by a high level of control on their generation and are somewhat flexible. This is opposed to *variable capacity* generators which depend on external factors such as weather conditions. Table 8.1, in Appendix 2.1, lists the common energy generation types as well as their intrinsic characteristics. In Belgium, the largest generator is *Engie Electrabel* [16]. To address the security of supply, the generation capacity must be correctly scaled. In this way, non-flexible generators are generally used to meet the base-load demand, while flexible generators are preferred to meet peaks in demand. Furthermore, a set of *reserves*, ranked according to their reactivity, is deployed in the network to help balance supply and demand.

## 1.2 Transmission Network

This is the part of the network responsible for long-distance transmissions. High-voltages (220kV-1000kV) are used to reduce transmission losses, even though HVDC starts to emerge too[1]. Due to the colossal operation cost, it is managed, in Europe, by a very limited number of operators, called the *Transmission System Operators*, which are organised in the *European Network of Transmission System Operators*, abbreviated ENTSO-E from now on. There is only one TSO in Belgium: *Elia* [19]. Transmission System Operators are also responsible for maintaining the instantaneous supply and offtake balance in the network.

## 1.3 Distribution network

The network distribution acts on a smaller geographical scale and brings access to electricity to consumers via medium and low voltage connections. It is managed by *Distribution System Operators* (DSO). In Belgium, we count several DSO who are dispatched on a geographical basis. ORES, Tecteo, Régie de Wavre, AIESH and AIEG are dispatched in Wallonia, Sibelga in the Brussels-Capital region, and Eandis and Infrax in Flanders.[2] They are also in charge of installing the billing metering device and performing the invoicing. The complete list can be accessed via [20].

## 1.4 Balancing supply and demand

In addition to the management of generation plants and their reserves, balancing *injections* (supply) and *offtakes* (demand) can also be tackled via the use of energy storage solutions. However, it is an expensive process which still represents a very difficult challenge to solve, even though many technologies have already been developed (see the article [21] for an overview). An alternative is the flexible *demand-response* approach, which strives to reduce demand in time of *scarcity*. This later mechanism can be enforced by developing market incentives such as *time-variant* pricing (*scarcity pricing*). Coupled with new infrastructures such as *smart grids* and *smart meters*, it is a source of promising improvement for the EU network [22].

## 1.5 Trends in the Belgian production, consumption and energy mix

Looking at the state of the Belgian electrical system reveals how strongly it is evolving with time. We encourage the reader to have a look at the Febeg 2018 report[23], written in collaboration with *Elia*, which integrates many relevant statistics about the Belgian electrical system usage. The following summarizes some of their key statistics for 2018.

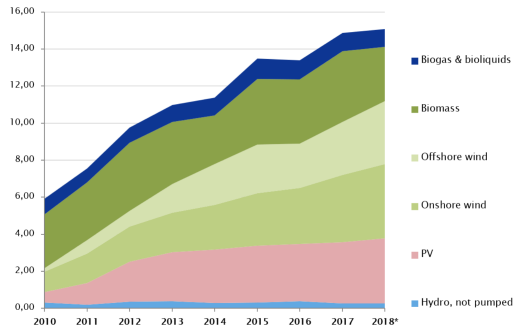
**Supply side** The net total electricity production in Belgium amounts to 69.2 TWh which has decreased by 24.4% since its peak of 91.5 TWh in 2010. This decrease is mainly explained by the unavailability of a part

of the national nuclear park. According to [24], imports have reached 8% of the electricity mix in 2017 but soared 22% in 2018, thus complexifying the analysis of the share of energy generation by types.

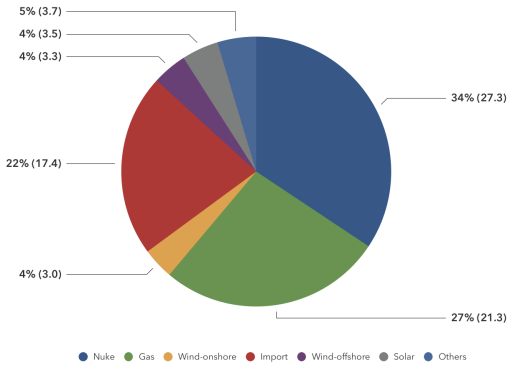
**Demand side** For the last 8 years, the total consumption remained stable but slightly decreasing, reaching 81.35TWh. This implies that the country has had a constant deficit in its production, which it compensates by importing the remaining from its neighbors. In 2017, the industry, services and residentials respectively represented 46.7%, 25.8% and 21.7% of the demand.

**Energy mix** The national installed *capacity* increased to 23.289 MW in 2018, of which 36% are considered renewable energy sources (RES). However, since the load factor (all the technologies considered) decreased as well, the total production declined. Still, the renewable energy production progressed by 18% from 2017 to 2018 making up almost a quarter of the total share of the 2018 production. Thus, three quarters of the production still originate from classical generation units.

Figure (2.2) depicts the state of the Belgian energy mix as of 2018. A detailed real-time monitoring platform can also be consulted at this web address [25].




(a) Evolution of Belgian renewable energy production in MWh [23].




(b) Belgian electricity mix along with imports in 2018 [% (TWh)]. Figures from [24].

Figure 2.2 – State of the Belgian energy mix.

# 2 Current electricity markets design

Since electricity is a commodity which can't be stored at large scale, different types of electricity markets have been arranged in a sequential order, starting years before the actual delivery and ending after the actual delivery [18]. Subsection 2.1 first highlights the recent changes in the EU market design as well as the market actors. Then, we give an overview of the most common electricity markets classified by types, end-consumer, geography and time scales.

## 2.1 Liberalization of energy markets

Up to three decades ago, the European energy market design was organized in a regulated monopoly. In this way, one or a small number of *vertically integrated* companies used to have the full responsability for the generation, transport and distribution of energy at a national level[18]. However, the European Commission, aiming at an international EU market, has recently decided to design an unbundled structure by opening the competition to third parties. As a result, gas and electricity markets have been liberalized by means of three legislative packages adopted starting from the 1990s (1996), 2003 and in 2009 [1]. These European directives are then being translated by the Member States but this long and complex modernization procedure is still ongoing as pointed out in their Energy Union 2015 report [26]. The flow diagram from Figure (2.3) summarizes the current Belgian situation from the electricity generation phase up to the retail. Missing notions will be shortly introduced in the following subsections.

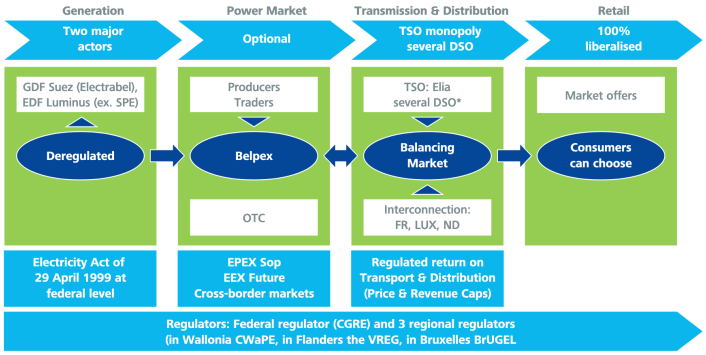



Figure 2.3 – Belgian market liberalization and organization flow diagram[2].

## 2.2 Retail and wholesale electricity markets

Several types of market coexist depending on the type of consumers, time scale and geography. By classifying the markets on the basis of their consumers, we can differentiate the *retail* market from the *wholesale* market. This distinction is motivated by the fact that residential needs differ from industrial ones. Indeed, it is easy to understand that small residential consumers would prefer paying a flat rate, be it, on average, a bit higher via an intermediate retailer, for their yearly electricity demand. Table 2.1 summarizes the main difference between the two.

| **Characteristics** | **Retail market** | **Wholesale market** |
|---|---|---|
| Consumer type | Small residential and industrial | Large industrial |
| Geographical scope | Local | Local and transnational |
| Actors | Suppliers, consumers | Generators, suppliers, industries |
| Types of contract | Tarif (periodic) | Many (derivatives, auctions, etc.) |
| Entrance fee | No | Yes |
| Price variance | Low | High |

Table 2.1 – Highlights of key differences between the retail and wholesale electricity market.

Market participants in the wholesale electricity market naturally involve *generators*, *Transmission System Operators* and *Distribution System Operators* (together called system operators) are paid for generating electricity, the long-distance transport of electricity (and its stability), and the delivery to *consumers* respectively. In an unbundled structure, it should be pointed out that DSO may not sell electricity to consumers. We designate as *retailers* any agent who buys electricity from generators on the electricity market and sells it to consumers not participating in it. To facilitate the interaction between the above mentioned economic agents, a *market operator* (MO) is responsible for issuing contracts between the supply and demand side according to the market mechanism. Finally, a *regulator* is involved to set the rules and oversees the functioning of the market. In Europe, ACER sets the guidelines for transnational electricity network and markets with the network code[27]. The federal Belgian regulator is the Commission for Electricity and Gas Regulation (CREG)[11]. A schematic simplified diagram of their interactions is available in Appendix 2.1.1 (Figure (**??**)).

## 2.3 Types of wholesale electricity markets

We can highlight three common types of wholesale electricity markets: power exchange, over-the-counter (OTC) and an organized OTC.

**Power exchange** A *Power exchange* market is a multilateral trading platform featuring anonymous and transparent pricing of standardized products. The market operator aggregates demand and supply bids of market participants and then perform a market clearing once per predefined time period in order to output a single market price.

**OTC** In contrast, *OTC* markets are used for bilateral trading and allow for the trading of assets without the formal structure of an official exchange platform. Prices of various products (custom time periods, blocks, etc.) are confidential and not transparent to other market parties, even though the market prices from the transparent Power Exchanges are mostly used as a reference.

**Organized OTC** Finally, *organized OTC* markets allow market participants to submit bids to a market platform in a similar manner to Power Exchanges with the exception these markets are continuously cleared. In other words, one market actor can accept an offer or a bid bilaterally, thus resulting in different trade prices [18].

## 2.4 Overview of the consecutive wholesale electricity market

Due to the technical difficulty and cost of storing electricity, contracts (transactions) are based on delivery at some time in the future which motivates the distinction and design of markets with different time horizons. Figure (2.4) represents of a schematic view of the markets discussed in this section and how they relate to one another in terms of time of delivery. Strictly speaking, day-ahead and intraday markets are forward/future markets since they gather actors interested in purchasing electricity for a future delivery but the denomination of *forward* markets tends to be reserved for time-horizons that take place before the day-ahead market. The imbalance is considered outside the spot market since it is undergone by the market actors.

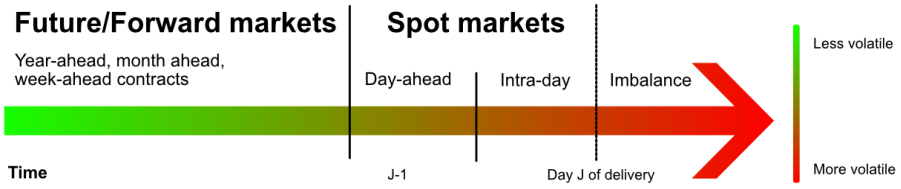



Figure 2.4 – Markets by time horizon

### 2.4.1 Forward and future markets

Trading in forward and future markets is done from years before the delivery period up to the day before the later begins. Thus, contracts are settled in such a way that the price for a quantity of electricity to be delivered in the future is agreed upon today. Its price should reflect the consensus expectation of the spot market and is used to reduce exposure to the volatility of the spot market. Indeed, large industrial consumers prefer to secure their future delivery consumption at an upfront known cost and reduce their exposure to unknown possible future sudden price increase[18].

**Forward market** A (firm) forward contract specifies the quantity, date of delivery, date of payment following delivery and the penalties if either party fails to honor its commitment [28]. Such type of contracts are usually non-standardized and traded bilaterally over-the counter [18]. Contracts can be settled on the basis of several time horizons. Year-ahead, quater-ahead and week-ahead are the most frequent types of time horizons encountered. In Belgium, long-term contracts are negociated on the ICE Endex[12] and the European Energy Exchange (EEX)[29]. Later, we refer as hedging the risk-reducing strategy consisting in purchasing on those markets with the objective of reducing the expose to the volatile prices of the *spot market*.

**Future market** In a future market, participants can buy and sell standardized (firm) forward contracts, which are, in opposition to the above, not backed by physical delivery. These contracts can then be further traded on power exchanges which is not usually the case for forward contracts [18].

### 2.4.2 Spot market

As the date of delivery approaches, participants must balance their position if they cannot produce, consume or store the energy. They might have also a better knowledge of their demand on a finer time scale (e.g. hourly) or leverage more accurate weather forecasts. The spot market is thus used to refine the traders' positions close to delivery time. However, most of the forecasting errors also end up in the spot market so it is known to be to be very volatile (even down to negative prices) and risky which is why smaller amounts are usually traded. Under the spot market denomination, we distinguish the *day-ahead* and *intra-day* markets.

**Day-ahead market (DAM)** Electricity for next-days delivery can be traded bilaterally (OTC trading) or on the day-ahead power exchange for each hour of the next day. In the Belgian market zone, the power exchange platform is called Belpex DAM and is managed by EPEX SPOT Belgium. The DAM power exchange is said to be a fixed gate auction principle which means that final electricity prices are *cleared* (published publicly) at a fixed specified time: the gate closure.

**Intra-day** Trading in the intra-day market happens on the delivery day itself and is used to correct shifts in the day-ahead nominations due to better forecast or unplanned outage[18]. Market actors submit their orders for the following hours of the trading/delivery day. The Belpex Continuous Intra-day Market is an organized OTC market that relies on a continuous gate auction mechanism. In other words, direct anonymous contracts are settled as soon as a deal is feasible: the market is cleared continuously.

**Balancing market** The balancing market is designed to deal with real-time imbalance in the grid and not for direct trading of electricty since it is undergone by the market participants. In this way, the imbalance price is given by a function of the reserve activated by the TSO to maintain the balance of its system.

## 2.5 Market zones

Finally, markets may also vary depending on their geographical scope. Several trading platforms, or *market zones* exist within the EU territory. The European energy legislation strives to design, in the long run, increasingly interconnected market zones with, as a result, prices converging to the same level. This is also known as *Market coupling*[23]. Forward and future contracts can be traded within a market zone or between them but for cross-border allocation. Transmission capacity is traded apart from the energy. The day-ahead market coupling is growing across Europe as pointed out by TSOs and regulators[30]. The current state of the Power exchanges interconnections in Europe as well as an highlight of the price convergence are exposed in Figure (8.4) and Figure (8.5a) respectively in the Appendix 2.1.3.

# 3 Trading in power exchange markets

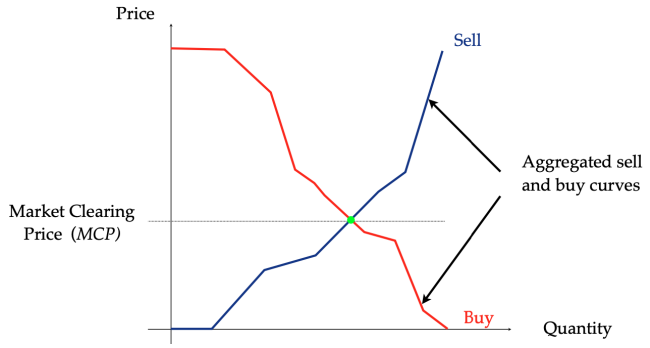
The following covers some of the mechanics and dynamics of auction based electricity pools.

Since energy is pooled from generators on its way to the load, the concept of centralized electricity pool has been introduced to provide a mechanism for reaching a market equilibrium. For the same reason, different forms of bilateral trading exist depending on the amount of energy to be traded and the time available. In the later, we refer by trade an amount of megawatts-hours to be delivered over a specific period of time [28].

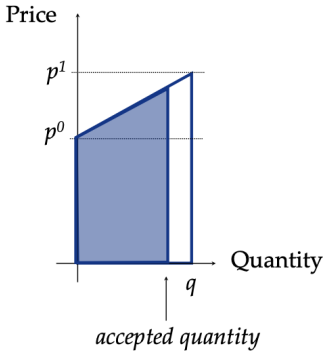
## 3.1 Notion of fixed gate auction mechanism

Generators submit *bids*, an amount of electricity at a certain price, for the period under consideration. Consumers submit *offers*, an amount of electricity they are willing to pay at a certain price. At gate closure, the market is cleared, i.e. an algorithm determines which order should be accepted and at which price. To do so, bids are ranked in ascending order based on their price (*merit order*) and a *supply curve* is drawn. In a similar manner, a *demand curve* is built in descending order of the offers' prices. Both of them can be schematically displayed as in Figure (2.5a) and their intersection marks the Market Clearing Price (MCP), also called the System Marginal Price or *equilibrium price*. The market operator then accepts *in-the-money* transactions, i.e. any bids inferior to the MCP and offers above the MCP. Generators thus receive the price corresponding to offers which satisfy the marginal demand. In Figure (2.5a), this means that every operations at the left of the green dot are accepted and paid at the MCP. Ideally, *at-the-money* orders, like depicted in Figure (2.5b), should be accepted in proportion $\frac{MCP-p_0}{p_1-p_0}$ if $p_0 \neq p_1$. However, since GenCos want to recover their start-up costs and consumers secure their supply with confidence, *block order*s are used in practice [9]. In other words, *bids* and *offers* are considered as non-breakable blocks. For the market participants, this means that their orders are either fully rejected or fully accepted at gate closure.

The application of this mechanism has led to the so-called *merit order effect*, whereby European prices tend to decrease as a result of the growing capacity of variable generators (e.g. wind, hydro and solar) in the grid[1]. Since such generators have no fuel cost, they are more prone to offer lower prices and, therefore, increasingly influence the prices levels for classical generators.




(a) Establishement of the MCP [9].




(b) At-the-money transactions [9].

Figure 2.5 – Market clearing mechanism

## 3.2 Comparison of year-ahead and day-ahead markets

### 3.2.1 Belgian Day ahead market (DAM)

To order day-ahead quantities at auction, market participants submit their orders on a hourly basis in an order book that is cleared at 12:00 AM[18]. Supply and demand curves are thus computed for each hour slot of the next day. This mechanism implies that participants trade electricity based on the public hourly price distribution of the day before and receive confirmation or rejection for their orders, along with the hourly clearing price, only after the gate closure. The DAM covers all the hours of the day and is open throughout the whole year.

Figure (2.6) displays the full day-ahead data that have been fetched from the European Network of Transmission System Operators open platform [31] via their REST api. The historical data starts on the 1st of May 2015 at 2AM up to real-time with prices on a per-hour basis. The interactive visualization is done with the the Python module *plotly*. The plot confirms our previous statement, whereby the day-ahead prices are indeed very volatile. For example, on the 14th of December 2016, prices reached a soaring 696€/MWh at 5PM before dropping to 148€/MWh for the next hour slot. We also observe that prices can sometimes drop

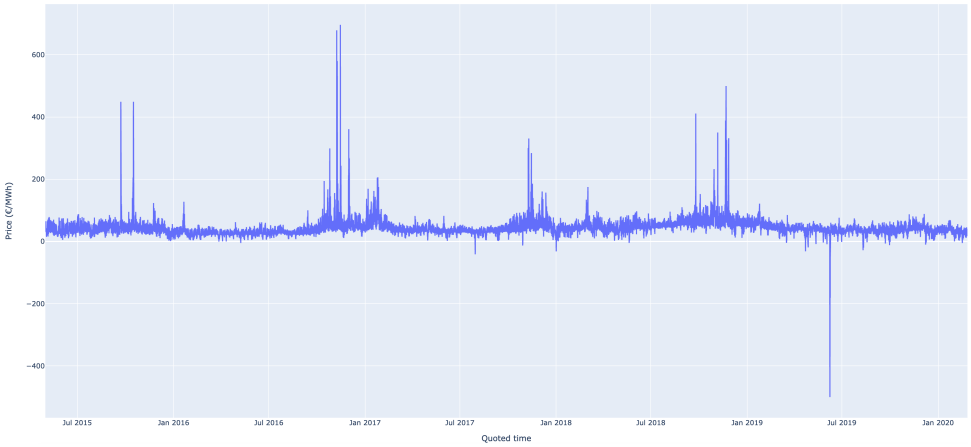



Figure 2.6 – Belpex day-ahead (DAM) market prices in €/MWh. Data fetched from the ENTSO-E REST api.

below zero with some huge spikes in some rare and extreme situations when supply is in excess.

### 3.2.2 Belgian year-ahead market (YAM)

Quotation time in the year-ahead market already starts 3 years before the year of delivery and ends the day before the start of the year of delivery. Prices are generally aligned on the ICE Endex[12]. Another difference with the DAM to highlight is that the YAM is only open during *working days*. Furthermore, quantities are not traded on a per hour basis but instead on a per day basis. Contracts that are signed on power exchanges agree on a certain electricity quantity to be delivered at a constant rate during the delivery year. We only cover the case of a *click-by-click* contract type in Subsection 3.4.

Figure (2.7) displays the year-ahead market prices at quotation time. The prices corresponding to the delivery year $y$ are evaluated by the quotation of the CAL $y$ which indeed starts three years before the delivery. This way, it is important to understand that during the three-year quotation period, a trader buying electricity when the quotation price is low would make a significantly better decision than another trader who would buy at times of higher prices since the product delivered is exactly the same and relates to the same delivery period. Contrary to the DAM, YAM prices never dropped below zero.


Figure 2.7 – Year-ahead market prices at quotation time.

### 3.2.3 Comparison of day-ahead and year-ahead prices

To compare the statistical properties of the day-ahead and year-ahead markets, it wouldn't make much sense to compare them for the exact same quotation time. Instead, statistical properties of the evolution of electricity prices should be derived on the basis of the delivery period (year) so that somewhat comparable products are analyzed. Following this observation, Table 2.2 presents the mean and variance of the DAM and YAM market prices for a common delivery period spanning from 2016 to 2019 both inclusive. This time range corresponds to the largest common delivery period between the acquired day-ahead and year-ahead prices datasets.

| **Market** | **Mean**[€/MWh] | **Variance**[(€/MWh)$^2$] |
|---|---|---|
| Day-ahead | 43.950 | 525.849 |
| Year-ahead (CAL 2016-2019) | 39.730 | 48.472 |

Table 2.2 – Mean and variance of historical year-ahead and day-ahead market prices for delivery between 2016 and 2019 both inclusive.

We first notice that the DAM has an approximately 10% higher average price than the YAM. As expected, its variance is also an order of magnitude higher than the one of the year-ahead market. In fact, had we skipped CAL 2019, the variance of the year-ahead market would even have been halved to 27.81(€/MWh)$^2$.

To illustrate the significantly higher *volatility* of the day-ahead market, it is worth comparing visually the two market prices for a common delivery year. For that purpose, Figure (2.8) concatenates their respective quotation periods for a delivery in 2016. The difference in volatility becomes striking to the eye.

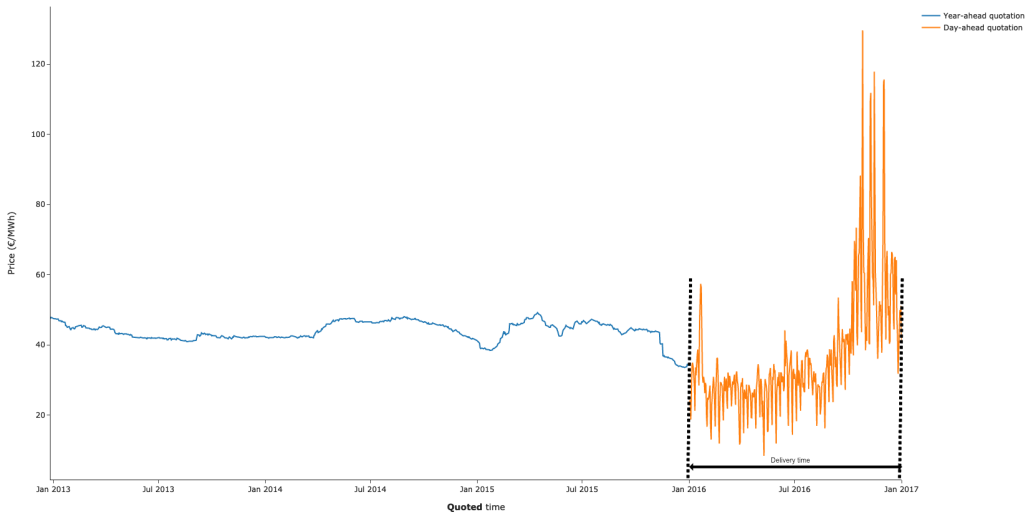



Figure 2.8 – Historical quotation prices for CAL 2016 and the daily averaged price of the DAM for the common delivery year 2016.

It would be worth investigating whether the year-ahead market is indeed consistently cheaper than the day-ahead market though. By refining our analysis on a per-year basis, we obtain the results of Figure (2.9). A similar plot (Figure (8.6)) from the official CREG report comparing different forward markets is also presented in the appendix Appendix 3.0.1. It can be deduced that the forward market is not always cheaper than the DAM. In 2015 and 2016, the DAM was actually cheaper than the forward market. It is noticeable that, in 2019 (CAL 2019), the forward market was characterized by a steep increase of the price (see Figure (2.7)), thus resulting in a high variance for this particular year. This suggests that trading solely in the forward market might not be an optimal solution and might not always prevent the exposure to spiky prices. *Arbitrage* is a trading strategy that consists in exploiting the market imperfections by purchasing some quantity in a first market and exploiting the price difference in a second one to make benefits. According to our observations, arbitrage strategies between the year-ahead and day-ahead market, i.e. selling exceeding base supply

purchased on the YAM on the DAM, may be profitable when performed at strategic times. Such strategies may however be hindered by transactions costs.

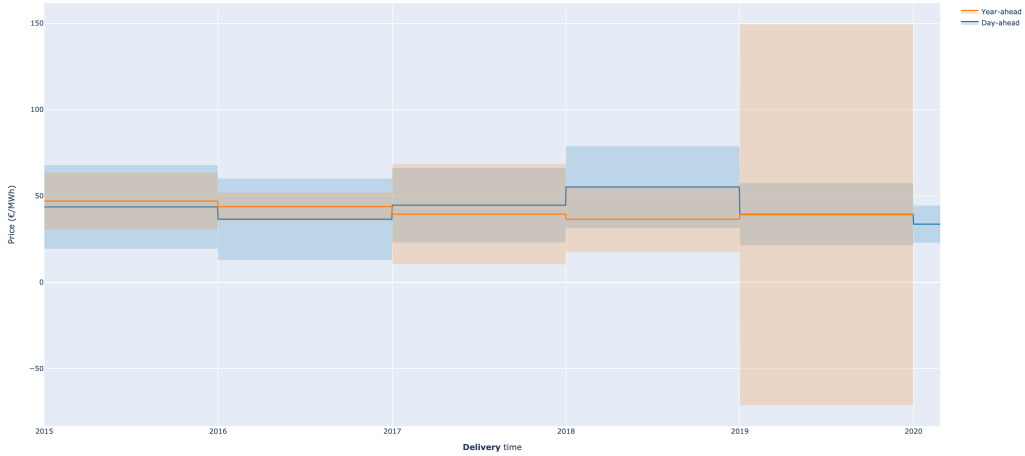



Figure 2.9 – Day-ahead and year-ahead electricity market prices at corresponding *delivery* time (€/MWh).

## 3.3 Power price drivers

Figure (2.10) summarizes a set of external factors that are known to have an influence on electrical energy prices. Those factors are briefly discussed in the following paragraphs.

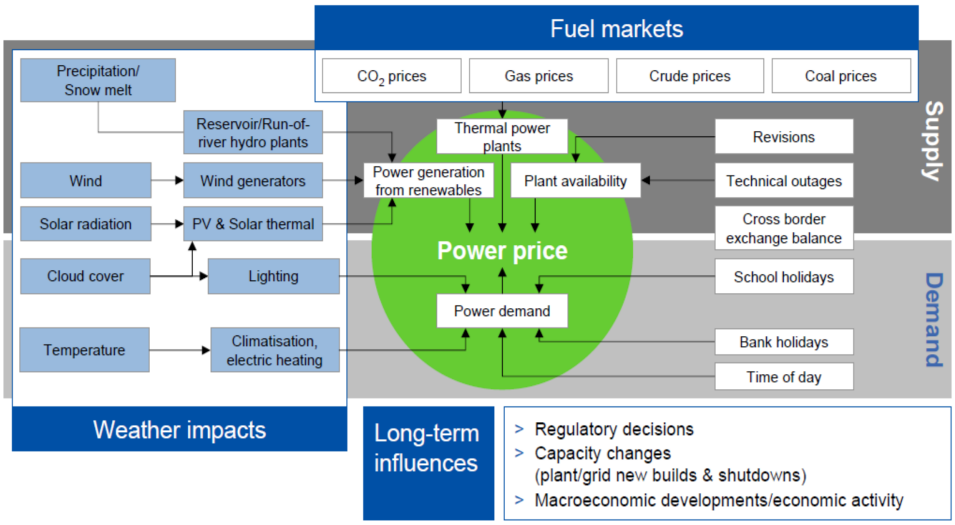



Figure 2.10 – Power price drivers [3].

**Weather conditions and renewable GenCo** With the increasing share of renewable generators in the Belgian electricity-mix (see Subsection 1.5), the weather is expected to have more impact on short-term electricity prices than ever before[28]. In this way, favorable weather conditions, like windy or sunny days, are likely to move prices downwards since renewable generators, which benefit from virtually free energy, would be ranked higher from a *merit order* perspective (Figure (8.3), Appendix 2.1.2). Another perspective is that consumers demand is naturally impacted by weather, especially temperature changes. Thus, the increasing capacity of renewable-based generation in the country, coupled with the unstability of weather conditions can increase the volatility of the prices while complexifying the management of traditional and less responsive generation companies. While the impact of weather is indisputable for short-term electricity markets, the later might not apply to the same extent for long-term markets.

**Regulation and capacity changes** It comes to no surprise that changes in regulations like the liberalization of the industry, changes in tax regimes or subsidies, among others, have a very strong impact on the energy market[32]. Those may however be tough to predict precisely in advance. An illustration of another change in regulation is the one related to carbon emissions which have set a penalty on the ton of $C0_2$. As pointed out in the 2018 Febeg annual report[23], the price per ton of $CO_2$ has almost quadrupled between the beginning of 2017 and the end of 2018. Such changes naturally impact the market dynamics.

**Capacity changes** Planned revisions or technical outages are illustrations of factors that influence the price of the energy from the supply side. For instance, a deficit in production due to a nuclear park under maintenance would likely require an increase of net imports.

Other factors of price fluctuations are periodic events like working days, time of the day or holidays as well as economical conditions (e.g. crise, increasing inflation, etc.) which have a natural impact on electricity demand. Other (alternative) energy sources like coal, crude oil and natural gas, just to name a few, influence each others which yet complexifies the analysis of electricity price fluctuations. Finally, the role of interconnection should be considered. For example, Germany's new strategy to use gas and coal instead of traditional nuclear plants tends to turn the country into a frequent exporter which can have an impact to some extent on neighboring prices.

From a forecasting perspective, the complex interactions between all these external factors make a manual evaluation of the future price movement far from trivial. A good expertise in the field is thus required for corporate energy managers. Yet, professional solutions leveraging these external factors have been developed to help these managers in their decision making[33, 34].

## 3.4 Click-by-click forward contracts

This subsection gives an overview of a generic contract details that could be settled between a small industrial company and its broker intermediate to trade in the wholesale electricity market via a virtual platform. The following covers a click-by-click contract type and will be used as a reference for this work.

### 3.4.1 Price terms

We refer by *Booked Power* the aggregated quantities registered by the customer via *click orders*. There are two pricing systems the customer can usually choose from. The first one is based on the ICE Endex[12] while the second one is determined on the basis of OTC ASK or BID quotations. The company books its *base load supply* on a yearly, quaterly or montly basis via either *Endex* or *OTC* orders. The difference between the customer's base load and its *nominated requirements* is called the *spot supply* and is expressed in MWh. Naturally, any future *imbalance* at delivery time is charged to the customer at tariffs set by the transmission system operator, like *Elia*[19] in Belgium.

### 3.4.2 Booking of forward power blocks

In practice, the digitalization of the economy has often pushed companies to automatically settle their contracts by means of a web platform. In this way, the customer can either click or declick *fixed-size* blocks of power for the desired delivery period on the price published by the ENDEX or use an OTC booking mechanism. Hence, the customer can build his contract portfolio in a somewhat flexible way like listed in Table 2.3. Endex orders are based on the end-of-day quotation principle. This means that the customer places an order based on the published price of the previous day without knowing the current price and only receives the billing invoice the following day. On the other hand, an OTC buy (click) order needs to be performed on another OTC ASK quotation. Conversely, an OTC sell (declick) order needs to be on an OTC BID quotation. Obviously, transaction fees may apply for both systems and may even vary depending on the booked delivery period. Fees might be asymmetric with respect to click or declick orders too even though it is not captured in the simplified notation.

| Price reference | Operation | Volume[MW] | Period | Transaction fee [€] |
|---|---|---|---|---|
| Endex | Click/Declick | k×block_size | CAL | $\texttt{fee}_{cal}$ |
| | | | Q | $\texttt{fee}_{Q}$ |
| | | | M | $\texttt{fee}_{M}$ |
| OTC | Click/Declick | k×block_size | CAL | $\texttt{fee}'_{cal}$ |
| | | | Q | $\texttt{fee}'_{Q}$ |
| | | | M | $\texttt{fee}'_{M}$ |

Table 2.3 – Typical click-by-click booking mechanism assuming yearly, quarterly and monthly ahead delivery options. $k$ designates the number of power blocks purchased.

With this booking mechanism, the monthly price paid by the customer for the active energy is given by:

$$\begin{aligned}\textbf{Bill for month } M_{bill} = &\sum_{period=\{M,Q,CAL\}} \sum_{i=1}^{n^{click}_{period}} \texttt{block\_size} \times (\mathrm{p}^{click}_{period,i} + \mathrm{fee}_{period,i}) \times \mathrm{H}_{period,M_{Bill}} \\ &- \sum_{period=\{M,Q,CAL\}} \sum_{j=1}^{n^{declick}_{period}} \texttt{block\_size} \times (\mathrm{p}^{declick}_{period,i} - \mathrm{fee}_{period,j}) \times \mathrm{H}_{period,M_{Bill}}\end{aligned} \tag{2.1}$$

where $\mathrm{H}_{period,M_{bill}}$ is the total number of hours of a given $period \in \{M,Q,CAL\}$ overlapping the billing month $M_{bill}$ . $p_{period,i}$ designates the price paid for the order triggered by click $i$ for the market (of $period \in \{M,Q,CAL\}$) under consideration.

Finally, additional costs inherent to the customer's spot supply naturally apply. For instance, in the case where the customer's *spot supply* is positive, i.e. when the base load is lower than his nominal requirements, the customer would be charged for the hourly settlement price (€/MWh) published by Belpex with an additional fee $\texttt{Fee}_{imbalance}$€. Conversely, for a *negative* spot supply, the customer is paid back but with a negative fee.

### 3.4.3 Constraints on the booking mechanism

A natural constraint of click-by-click contracts is that the broker agent enforces a minimum `block_size` order expressed in MW which can't be cut in smaller chunks. Additional possible constraints include a fixed deadline after which electricity can't be purchased anymore (close to delivery time) or even a limit on the cumulative monthly orders. If we designate by $\mathrm{Q}_{booked,M_{bill}} = \mathrm{Q}_{clicked,M_{bill}} - \mathrm{Q}_{declicked,M_{bill}}$ the booked power for month $M_{bill}$, the constraints can be generically expressed as:

$$0 \leq \mathrm{Q}_{booked,M_{bill}} \leq U_{M_{bill}} \tag{2.2}$$

$$\mathrm{Q}_{clicked,M_{bill}} \leq U_{M_{bill}} \tag{2.3}$$

$$\mathrm{Q}_{declicked,M_{bill}} \leq \mathrm{Q}_{clicked,M_{bill}} \tag{2.4}$$

$$\mathrm{Q}_{declicked,M_{bill}} \leq L_{M_{bill}} \tag{2.5}$$

# 3 | Problem Statement

This chapter frames the problem already introduced in Chapter 1 in a more formal way. First, we expose the data at hand as well as the set of constraints considered for the electricity booking mechanism. We derive from it a number of possible orders to place during the quotation period. From the combination of the above, we then identify that only the year-ahead market can be considered and introduce the time notation required to frame the problem, while assigning responsibilities between the operator and the solution that should be provided. By taking into account the customer's requirements, we finally propose to cast the problem as a probabilistic forecasting problem that can be extended to derive an index of opportunity.

## 1 Context formulation and notation

We consider the case of small generic industrial consumer which is interested in hedging against risk by a major part (e.g. 80%) of his total required electricity supply in the forward market and reducing the cost of electricity acquisition. From the discussion of Subsection 3.2, we can conclude that trading solely on the spot market should be excluded from the beginning. Strategically placing the acquisition orders of this proportion under controlled risk is the subject that interests us. Only the *year-ahead* wholesale electricity market is considered.

### 1.1 Risk, forecast and uncertainty

Reducing the cost of electricity acquisition by providing the right timing to place purchase decisions is a complex trade-off between strictly reducing risk and seeking opportunities which could lead to potential savings. *Risk* is captured in the *uncertainty* about the yet-unobserved price evolution and can lead to situations in which the company is constrained to buy electricity at any (high) price to meet its supply needs. Likewise, securing a position while prices are thought to be high while they are likely to decrease in a somewhat near future can also result in large opportunity losses or missed savings.

In the absence of any possible other contract types or sources to secure the electricity demand, relying on contract diversification is not possible. In this way, to strictly *hedge* against risk, i.e. to reduce the exposure to the highest prices, a baseline method would consist in dispatching uniformly as many purchase orders during the quotation period. This strategy strives to drive the average cost of electricity acquisition for a given CAL as close as possible to its mean price in expectation. For this baseline, the average acquisition price tends to converge to the average price for the complete quotation period, provided the allowed number of decisions tends to infinity too.

This strategy is however completely blind to the price evolution and ignores, additionally, possible predictable long or short term price movements that would result in potential large saving opportunities. Reducing the future price uncertainty would mitigate this problem. A look in this direction is thus needed.

### 1.2 Booking mechanism, data and constraints taken into consideration

**Market data** Historical data of the forward market is notoriously difficult to get one's hands on. The Elexys platform[35] gives only condensed information about the historical ICE Endex prices but a historical dataset

of the *year-ahead* market ranging from 2nd of September 2004 up to 16th of October 2019, is available for this Master thesis. It naturally features one price per market open-days. Figure (2.7) already exposed in a visual manner the available dataset.

**Booking mechanism** We consider the booking mechanism for the purchase of power assets to be the one of a typical year-ahead click-by-click contract type like exposed in Subsubsection 3.4.3. The minimum block power size is fixed to either 1MW or 0.2MW.

**Customer's demand** The company has a limited number of orders to place during the quotation period due to the comparatively large minimum block power size imposed by the click-by-click contract. The typical base demand to secure on the year-ahead market would be around $10MWh$ (to be delivered at a constant rate for one year). By considering the minimum `block_size` imposed by the click-by-click contract of 1MW or 0.2MW, we derive directly a maximum number of orders to dispatch for 3 years equal to 10 and 50 respectively.

### 1.3 Year-ahead market trading

In a forward trading context, we naturally highlight two main time periods. We designate by $\tau_Q^y := \{1,\dots,T_Q^y\}$ the price quotation (Q) period for CAL $y$ divided in $|T_Q^y|$ time intervals. Likewise, $\tau_D^y$ is the corresponding delivery (D) period. Since the methodology from trading in one CAL to another is the same, we will loosely drop the upper index $y$ for convenience unless precising the CAL is relevant in the context. For each given CAL, the full period of consideration $\tau = [\tau_Q, \tau_D]$ thus spans over 4 years in total. The time interval resolution $\Delta_{t_Q}$ is equal to 1 day $\forall\, t_Q \in \tau_Q$.

A decision at each $t_Q \in \tau_Q^y$ (up to three years before delivery) should be taken for the delivery of electricity at a constant rate for all $t_D \in \tau_D^y$. The decision is under the responsibility of a (human) operator who can decide whether to buy or sell a multiple of `block_size` or simply to take no action. An operator's decision taken at a time $t_Q$ is denoted by $u_{t_Q} \in \mathcal{U}_{t_Q}$ where $\mathcal{U}_{t_Q}$ denotes the feasibility set of actions compatible with the contract constraints at time $t_Q$. Logically, making sure that the whole of the forecasted base load demand is purchased by the quotation time horizon $T_Q$ is thus also under the responsability of the operator.

When the time interval following $t_Q = T_Q$ is reached, i.e. when the quotation period is over ($t_D = 1$), the delivery starts and the operator is left with no decision to carry out anymore. This thus motivates to recenter the time indexing in the quotation period only.

## 2 Decision stages with a Market Opportunity predictor

### 2.1 Decision stages

The evolution of the electricity price for a given CAL over a finite time horizon $T$ can be represented, in a discrete time model, by a random process $\xi = (\xi_1, \xi_2, \dots, \xi_T)$ with the dynamics of the process inferred from historical data. At a given present time $t \in \tau_Q$, the realization of the sequence of continuous random variables (r.v.) $\xi_{1:t-1} = (\xi_1, \dots, \xi_{t_0-1})$ is known and the residual uncertainty is represented by the probability distribution of $\xi_{t:T_Q} = (\xi_{t_0}, \dots, \xi_{T_Q})$ conditioned on the observed r.v. $\xi_{1:t-1}$ denoted $P(\xi_{t:T_Q}|\xi_{1:t-1})$.[13].

Table 3.1 summarizes the decision stages that should be undertaken by the operator. We consider the case of a *price-taking* consumer which assumes that the operator can purchase any quantity at the market price without affecting that price. The later is thus said to be *exogenous* to the consumer. This assumption seems legitimate due to the relatively small amounts traded by, for example, the SWDE, compared to the total volumes exchanged in the YAM and allows to simulate, in advance, possible realizations of the random process without caring about possible future actions of the operator.

| Stage ($t$) | Available information for taking decisions | | | | Decision |
|---|---|---|---|---|---|
| | Prior decisions | Observed prices | Residual uncertainty | Market indicator | |
| 1 | None | None | $P(\xi_1, \xi_2, \ldots, \xi_{T_Q})$ | $\hat{\mathscr{M}}_1$ | $u_1 \in \mathscr{U}_1$ |
| 2 | $u_1$ | $\xi_1$ | $P(\xi_2, \ldots, \xi_{T_Q} \mid \xi_1)$ | $\hat{\mathscr{M}}_2$ | $u_2 \in \mathscr{U}_2$ |
| 3 | $u_1, u_2$ | $\xi_1, \xi_2$ | $P(\xi_3, \ldots, \xi_T \mid \xi_1, \xi_2)$ | $\hat{\mathscr{M}}_3$ | $u_3 \in \mathscr{U}_3$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| T | $u_1, \ldots, u_{T-1}$ | $\xi_1, \ldots, \xi_{T-1}$ | $P(\xi_{T_Q} \mid \xi_1, \ldots, \xi_{T-1})$ | $\hat{\mathscr{M}}_T$ | $u_T \in \mathscr{U}_T$ |

Table 3.1 – Decision stages during quotation period (derived from [13])

## 2.2 A Market Opportunity Estimator (MOE)

Since the freedom of making decisions should be left to the operator, the problem lies in the design of a *Market Opportunity Estimator* $\hat{\mathscr{M}}_t$ that leverages historical prices data to output a single number, at time $t$ of evaluation, summarizing the confidence in pursuing a (single) positive click order which would lead to likely favorable money savings with a small worst case loss for a given forecast horizon (which could, for example, be the next hedge period of the baseline described in Subsection 1.1).

The forecast horizon $T_F \in \{t_0, \ldots, T_Q\}$ used for the evaluation of the *Market Opportunity Estimator* (MOE) at time $t = t_0 - 1$ is an hyper-parameter. The underlying forecast period is then denoted by $\tau_F := (t_0, \ldots, T_F)$.

Formally, a MOE for a given time $t$, with a chosen forecast horizon $T_F$, is a map from inputs $X_t \in \mathbb{R}^{m \times n}$ to a real output $\hat{\mathscr{M}}_t^{T_F} \in \mathbb{R}$. When only the historical (CAL y) price data up to time $t$ is used to produce the indicator, $X_t$ simplifies to the observed data $X_t = (\xi_1, \ldots, \xi_{t_0-1})$ but one could also embed additional features to improve the quality of the estimator. For example, additional categorical information, correlated time series or forecast of other commodities can be included in the information set. The output $\hat{\mathscr{M}}_t^{T_F}$ takes its values in $[0, 1]$ with 0 corresponding to a *wait* recommendation and a 1 to a positive **buy** feedback. A value of 1 would for instance correspond to the highest confidence of making savings by purchasing a `block_size` at evaluation time $t$ compared to any potential later purchase in $\tau_F$.

## 2.3 A two-stage framework

One approach to build a *Market Opportunity Estimator* is to split it into two subproblems corresponding to two sequential stages: a price forecast followed by a measure of the buying opportunity over the underlying predicted prices distribution. The reliability of the indicator is conditioned on the one of the forecast.

**Prediction interval forecasting** Measuring uncertainty is important to characterize and nuance the quality of a prediction and, obviously, to make correct informed decisions. This is why a forecast model $\mathscr{F}_\theta$ is used to leverage a set of observed values in order to predict, a probability distribution $\hat{P}(\xi_{t_0}, \ldots, \xi_{T_F} \mid \xi_1, \ldots, \xi_{t_0-1})$ of the future outcome conditioned on the previously observed prices. This refers to density forecasting. However, such ideal forecast can rarely be matched in reality, especially without making assumptions about the price data distribution. Simpler probability interval will thus also be considered. Choosing a model suitable for an accurate multi-step forecast of the price evolution $\hat{\xi}_{t_0:T_F}$ is a first crucial subproblem that should be tackled in order to produce, at a later stage, a good measure of opportunity. Forecasting with multivariate inputs is considered as an extension to this problem.

**Measure of opportunity** The second subproblem then consists in designing a *deterministic* mapping $\mathscr{M}_{t=t_0-1}^{T_F}$ from $\hat{P}(\xi_{t_0}, \ldots, \xi_{T_F} \mid \xi_1, \ldots, \xi_{t_0-1})$ to $[0, 1]$. We refer to this mapping as a *Market Opportunity Indicator* or opportunity measure, the goal of which being to capture the profitability and uncertainty associated to pursuing a positive click order under the current market conditions at time $t = t_0 - 1$[1] according to an

[1] compared to purchasing electricity during the interval spanning from $t_0$ to $T_F$.

estimator whose reliability has been correctly *backtested*.

Wrapping it altogether, we thus have to compute at evaluation time $t$: $\hat{\mathscr{M}}_t^{T_F} = \mathscr{M}_t^{T_F}(\mathscr{F}_\theta(X_t))$.

# 4 | Literature review

Due to the high complexity of the physical infrastructure as well as the large amounts of money traded in electricity markets, it comes to no surprise that the literature covering electricity market analysis, electricity contracts portfolio management and electricity forecasting is vast. Surprisingly though, the case of a small industrial consumer, with a risk-averse profile, looking to reducing its cost of electricity acquisition with only one type of derivative contract within the European wholesale market doesn't seem to be covered in literature, at least to the best of our knowledge. In such context, the topics covered in the literature bound to the financial electricity sector rather tends to focus on hedging through diversified contract optimization over different time scales[36, 37, 38], sometimes coupled with arbitrage strategies between markets[39, 40], or energy forecasting [41, 42, 43] but only for short or medium term delivery markets. All of which don't directly fit our particular problem focus.

At the crossroad between financial portfolio and supply chain management, optimization and forecasting, the well known *Value-at-Risk* (VaR) framework is central to risk management; a mathematical formulation of which is briefly presented in the Appendix 4. Without loss of generality, for a considered asset [1] , the VaR framework postulates the knowledge, or at least the estimation, of a certain probability distribution over its future outcomes. In this way, it provides a tool to evaluate the cost-benefits of investing in an asset, or a pool of such, and to improve risk-controlled decision making by focusing on the worst case outcome, i.e. the leftmost part of the distribution for a given confidence level $\alpha \in [0,1]$. In other words, VaR can be used to define hedging strategies for the construction of an optimized portfolio of diversified assets for an *a-priori* fixed acceptance risk level. Our problem instance differs a bit from this perspective since it leaves us with no other substitute contracts or markets to meet the electricity demand and, therefore, no "best" worst case scenario to select.

From the above observations, we thus propose to embrace a more generic approach centered around forecasting to deal with the *timing* component of the purchase decision process. More specifically, to reconciliate traditional forecasting techniques with the notion of *informed* decision making, we focus on *probabilistic* forecasting techniques.

## 1 Towards deep learning probabilistic forecasting

### 1.1 The probabilistic forecasting paradigm in an electricity and finance context

With the introduction of smart grids, renewable integration requirements and the modernization of the energy industry, probabilistic forecasting recently gained attention as market stakeholders started to face increasing uncertainty in their decision making processes[45]. It hasn't always been the case though. In a comprehensive review on Electricity Price Forecasting (EPF)methods published in 2014, Rafal Weron speculated probabilistic forecasting would develop in the next decade[41]. The paper classified and provided an overview of many popular techniques in econometrics and EPF but mostly applied to short term markets. While statistical models like autoregressive and exponential smoothing based techniques [46] are still popular benchmarks nowadays[47, 48], the recent advances in computational intelligence tool, like neural

[1] We simply denote by *assets* any economic value (e.g. contract, commodity product, stock) that an individual, corporation, or country owns or controls with the expectation that it will provide a future benefit[44]. For our problem instance, the benefit is to *meet the* future electricity *demand*, with the additional objective to do so in the most economical way.

networks, were already noticeable enough to provide successful results thanks to their ability to capture the non-linear and complex behavior of electricity prices[41].

In fact, *probabilistic* EPF truly gained momentum with the Global Energy Forecasting Competition of 2014 (GefCom2014[43]), which featured a price forecasting track. The objective was to predict 99 quantiles as a reasonable discrete approximation of the predictive distribution of the next 24 hours of short term electricity prices (which exposed naturally a non-negligeable seasonal component).

The $q$th *quantile*, or percentile, of a random variable Y is the value $y_q$ below which a fraction q of observations of this random variable falls. For instance, for a quantile q, $y_q$ is that value of Y which satisfies $F_Y(y_q) = q$ where $F_Y$ denotes the cumulative distribution of $Y$. The competition was designed such that assimilating the true distribution as the discretized forecast distribution (density forecast) would be an optimal strategy in expectation.

A simpler alternative to density forecasting is Prediction Interval forecasting. We denote by Prediction Interval (PI) an interval which contains the true values of future observations with a specified probability [41]. The later can either be predicted directly or derived from two symmetric quantiles if the estimator is able to produce the full future distribution. For instance, the $(1-\alpha\%)$ PI, centered in the median ($q_{0.5}$), features the interval extrema from the $\frac{\alpha}{2}\%$ and $1-\frac{\alpha}{2}\%$ quantiles with $\alpha \in [0,1]$.

The inclusion of PI forecasting entries, with a value of $\alpha = 0.05$, was considered in the renowned M4 competition [47, 49] back in 2018. Like GefCom2014, this competition positively influenced research by providing a common evaluation scheme and datasets to ease scientific comparison between methods[50]. However, unlike the previously mentioned one, it wasn't just bound to the electricity sector but, instead, featured more than 100 000 time series with various frequencies, origins (e.g. *Micro, Macro, Finance*, Industry, Demographic) and, therefore, dynamics[51]. Noticeable and valuable results are that forecasting accuracy of simple methods are not too far from those of accurate methods. Moreover, hybrid models which combine statistical with Machine Learning (ML) techniques yield the best results while pure ML methods tend to perform poorly[47].

In the absence of true seasonality in our data (see Subsection 2.2), the more recent results of the M4 competition are more valuable. We plan on evaluating our results with a similar approach to them (see Subsubsection 1.4.3).

## 1.2 Deep learning based time series modeling

While neural networks don't outperform statistical methods in a consistent manner[52] , time series modeling have mostly focused on *local* models whereby the parameters of the time series model are estimated per time series individually eventhough a collection of related time series may be available as pointed out by [53]. Recent advances in deep learning have however led to substantial improvements over this traditional approach by introducing *global* models that can leverage a large collection of time series data to estimate the parameters of a *single* model[4, 6, 54]. This later approach was also adopted by the winning hybrid model of the M4 competition[48]. [53] actually qualifies this family of models as *"having better capability to extract high-order features and identify complex patterns"* while [4] states it allows to fit more complex and potentially more accurate models without overfitting.[4, 6, 54]

In this context, the open-source Gluon-ts library [53], developed by Amazon scientists, was very recently introduced to fill the gap in the forecasting toolkit landscape by proposing the first deep learning library featuring state of the art *global* deep learning models and more conventional probabilistic and statistic models under a common (Python) interface. Notably, the library aims at rapid experimentation and benchmarking of *various* models with support for the creation of custom PI or density forecast models on top of the MXNET backend. In lights of the results discussed above, the Gluon-ts library seems to offer currently the best set of features to handle the complex dynamics ruling the year-ahead market prices. Hence, from a pragmatic point of view, the next section will only give an overview of the most important models featured in the library and which we plan to test.

## 2 An overview of the time series models implemented in Gluon-ts

We give now an overview of best performing models implemented in the Gluon-ts library and highlight their benefits for our comparison study. Naturally, to keep things clear and organized, we redirect to the original paper for more information regarding their details and implementation.

### 2.1 Notation and background

Let $Z = \{\boldsymbol{z}_{i,1:T_i}\}_{i=1}^{N}$ be a set of N univariate time series with $\boldsymbol{z}_{i,1:T_i} := (z_{i,1}, z_{i,2}, \ldots, z_{i,T_i})$, and $z_{i,t} \in \mathbb{R}$ corresponding to the value of the i-th time series at time $t$. The time point $t_0 = T_i + 1$ is referred to as the forecast start time and $\tau \in \mathbb{N}_{>0}$ to the forecast horizon yielding the last forecast step $T_F = t_0 + \tau$. As follows, $\{1, \ldots, T_i\}$ is the training (conditioning) range and $\{T_i + 1, \ldots, T_i + \tau\}$ is the prediction range. Let us also consider a set of time-varying covariate vectors $X = \{\boldsymbol{x}_{i,1:T_i+\tau}\}_{i=1}^{N}$ with $\boldsymbol{x}_{i,t} \in \mathbb{R}^D$. Most of the models discussed above support those *features* which might as well correspond to static attributes for each series or known future events. The inclusion of such features also allows the generation of "what if" scenario if so desired.

In the multivariate probabilistic forecasting paradigm, we define *forecasting* as predicting the probability distribution of future values $\boldsymbol{z}_{i,T_{i+1}:T_{i+\tau}}$ given the past values $\boldsymbol{z}_{i,1:T_i}$, the covariates $\boldsymbol{x}_{i,1:T_i+\tau}$ and the model parameter $\Phi$:

$$p\left(\boldsymbol{z}_{i,T_i+1:T_i+\tau} | \boldsymbol{z}_{i,1:T_i}, \boldsymbol{x}_{i,1:T_i+\tau}; \Phi\right) \tag{4.1}$$

We generically denote by $\Phi$ the set of parameters of the model which are learned jointly for all N similar time series.

**Autoregressive models** Autoregressive models[46] are a family of time series model which forecast the variable of interest based on a linear combination of predictors. In its simplest form, the predictor consists of lagged values of the variable of interest. An autoregressive model of order p takes the following form

$$z_{i,t} = c + \phi_1 z_{i,t-1} + \phi_2 z_{i,t-2} + \cdots + \phi_p z_{i,t-p} + \epsilon_t \tag{4.2}$$

whith $\epsilon_t$ being white noise. These models are trained on a sequence but reduces the prediction to a one-step-ahead problem which models $p\left(z_{i,T_i+1} | z_{i,1:T_i}, x_{i,1:T_i+\tau}; \Phi\right)$. In this sense, to generate multi-step prediction, they rely on a *recursive* scheme in which previous 1 step predictions are ingested as part of the true observations (training range) to produce the next step. Likewise, errors from the sequential one-step-ahead prediction are aggregated for the model update (which therefore leads to error accumulation too).

**RNN, LSTM and Seq2Seq** Recurrent Neural Networks[55] (RNNs) represent a class of artificial neural networks whose connections between nodes form a directed graph with the objective to learn a *fixed-length non-linear* representation of a sequence of arbitrary length[6]. Vanilla RNNs aim at capturing temporal dynamic in a time series but struggle to exploit long term dependencies (the so-called vanishing gradient problem). To cope precisely with this issue, LSTM (RNN) networks, relying on Long-Short-Term-Memory (LSTM) cells[56], have been introduced. Sequence-to-Sequence RNNs (Seq2Seq)[57] have finally been introduced to generate future sequences whose length doesn't especially equal the one of the input sequence. They are based on a encoder-decoder RNN architecture where the encoder generates a fixed-size summary context vector which is fed to the decoder.

### 2.2 Baseline algorithms

The library features several benchmarking univariate time series model through a Python interface of the R *forecast* package [58] which are worth reporting. Prominent models include two of the most widely used approaches to time series forecasting that are ARIMA[59] and exponential smoothing based models which

are discussed in the comprehensive ebook *Forecasting Principles and Practises*[46]. The library also includes an interface to the bayesian structural time series model *Prophet* developed by Facebook.

### 2.2.1 Prophet

Prophet[60] is an intuitive regression model with easily interpretable parameters which can be tailored according to the domain knowledge of the data at hand. It is specifically designed for business time series though. In this sense, the model is not the most adapted to our problem instance but would provide an interesting point of comparison. A time series $z_{i,1:T_i}$, abbreviated $z_t$, is modelled with three main additive components: trend ($g_t$), seasonality ($s_t$) and holidays ($h_t$).

$$z_t = g_t + s_t + h_t + \epsilon_t \tag{4.3}$$

*Trend* exits when there is a long term increase or decrease in the data. *Seasonality* , on the other hand, translates a periodic pattern that repeats with a fixed frequency. $h_t$ in Equation (4.3), represents the effect of holidays which occurs on potentially irregular schedules while $\epsilon_t$ is an error term which is not taken into account by the model and assumed to be normally distributed. The most relevant trend model in the paper simply considers a *linear trend* with a growth rate of $k$. Trend changes are then incorporated by defining so-called *change points* so that the piece-wise constant growth rate $k$ is allowed to change over time. In this way, this mechanism allows to split the time series into separate regimes with different dynamics. Let S be the number of change points (hyperparameter) dispatched over several timestamps $s_j = \{1, \dots, S\}$ and $\boldsymbol{\delta} \in \mathbb{R}^S$ be a rate adjustment vector. The rate at any time $t$ is given by the base rate plus all the adjustments encountered so far:

$$k_t = k + \boldsymbol{a}(t)^T.\boldsymbol{\delta} \text{ where } a_j(t) = \begin{cases} 1, & \text{if } t \geq s_j \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

If the change points are not explicitly defined, the selection can be done automatically by putting a sparse prior on the adjustment vector. For instance, the sampling is done according to a Laplace distribution such that $\delta_j \sim \text{Laplace}(0, \tau)$ where $\tau$ is used to control the flexibility of the model in alternating its rate. Future rate changes are then emulated by replacing $\tau$ with a variance inferred from data and sampled randomly by assuming the average frequency of change points matches the one of the historical data. Periodic effects are approximated by means of standard Fourrier series whose parameters are normally sampled past the observations. Holidays are handled in a similar way but with a vector of Indicator functions indicating whether the time $t$ belongs to a given holiday.

All in all, the modeling is intuitively apparent to a curve-fitting task, in opposition to autoregressive models which take into account the time dependencies in the data. Advantages of such approach over this alternative is faster fitting, no need for interpolation of missing values and straightforward interpretability at the cost of a probably reduced inference capacity, more arbitrary hyperparameters, and strong modelling assumptions though.

### 2.2.2 Non-Parametric Time Series forecaster (NPTS)

NPTS[61] is a *non-parametric* forecast method used by AWS which falls into the category of autoregressive forecast methods. Traditional naive forecasts would, for example, use a fixed past time index, like $T-1$ or $T-\tau$ as the prediction for the time step $T$ but, instead, NPTS samples randomly a time index $t \in \{0, \dots, T-1\}$. By doing so, one obtains a Monte Carlo Sample which allows NPTS to generate prediction interval from the predictive distribution. The generative process for time T is

$$\hat{z}_T = z_t, \quad t \sim q_T(t), \quad t \in \{0, \dots, T-1\} \tag{4.5}$$

where $q_T(t)$ is the (categorical) probability distribution over the indices in the training range. In this way, the model indeed reduces to an autoregressive model with weights $q_T(\cdot)$ since the expectation of the prediction yields:

$$\mathbb{E}[\hat{z}_T] = \sum_{t=0}^{T-1} q_T(t) z_t \tag{4.6}$$

Multi-step predictions is achieved by absorbing the predictions for the last time steps into the observed set and then generating predictions using the last T values. To sample more recent values in the observed set with higher probability, the sample distribution $q_T$ uses weights that decay in an exponential way according to the distance to the past observations:

$$q_T(t) = \frac{\exp(-\alpha|T-t|)}{\sum_{t'=0}^{T-1} \exp(-\alpha|T-t|)} \tag{4.7}$$

where $\alpha$ is an hyper-parameter that is adjusted based on the data. Letting $\alpha = 0$ would, for instance, result in uniform sampling while setting $\alpha \to \infty$ degenerates the prediction to the one of a naive random walk forecast which predicts the last observed value. This thus makes NPTS an interesting baseline algorithm to include in our study.

### 2.2.3 Feed Forward Neural Network (FFNN)

The library includes a last baseline algorithm called a Simple Feed Forward Neural Network. It takes an input window of size $W$ and learns to predict the distribution of the values of the subsequent future window of length $W$. The model first process the input window with a mean scaling before feeding it into a multi layer perceptron with a simple architecture. In this way, the model consists in forty dense layers with *ReLu* activation but don't include any dropout layers. The number of hidden nodes in each layer is also set to 40. The model computes a probability distribution for future data given the past observations and draws N samples from it to derive prediction quantiles. To do so, the network needs to estimate the parameters of the output distribution. For example, in the simple case where we assume a Gaussian distribution, the model need to learn the mean and the variance that fully specify the distribution. Since each distribution output have its own parameters that needs to be estimated, the FFNN first output an intermediate vector whose dimensions matches the prediction length and the number of features. Then, a projection layer is concatenated to the model to map its intermediate output to the parameters of the distribution. The model is then trained by optimizing the negative log-likelihood under the model parameters and he parameters of the projection layer are optimized along with the rest of the network. By default, Gluon-ts always considers a default t-student distribution and a decaying learning rate.

## 2.3 DeepAR

DeepAR[4] is probably the model which is the most highlighted in the library and AWS, which is unsurprising considering it is developed by Amazon scientists like the library. The model sets high expectations with good results achieved on demand forecasting tasks which notoriously violate core assumptions of simpler models like Gaussian noise, stationarity or Homoscedasticity[4]. They promise minimal manual feature and hyperparameter engineering to capture complex, group dependent behavior, all of which with minimal historical data. This should indeed set DeepAR apart from classical forecasting approach[41] and makes it a good candidate. Figure (4.1) gives a visual summary of the model.

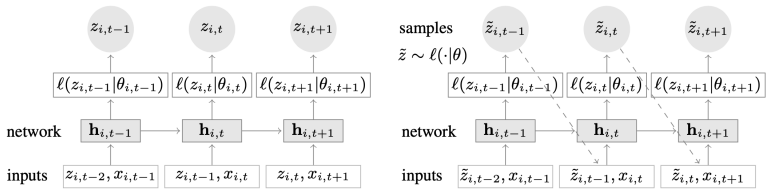



Figure 4.1 – Summary of DeepAR model. The left part depict the training phase and the right part the prediction phase. *Source*: [4]

The output model distribution, which they denote $Q_\Theta\left(\boldsymbol{z}_{i,t_0:T_F}|\boldsymbol{z}_{i,1:t_0-1},\boldsymbol{x}_{i,1:T_F}\right)$ consists of a product of likelihood factors:

$$Q_\Theta\left(\boldsymbol{z}_{i,t_0:T_F}|\boldsymbol{z}_{i,1:t_0-1},\boldsymbol{x}_{i,1:T_F}\right)=\prod_{t=t_0}^{T_F}Q_\Theta\left(z_{i,t}|\boldsymbol{z}_{i,1:t-1},\boldsymbol{x}_{i,1:T_F}\right)=\prod_{t=t_0}^{T_F}\ell\left(z_{i,t}|\theta\left(\boldsymbol{h}_{i,t},\Theta\right)\right) \tag{4.8}$$

parameterized by the output (hidden state) $\boldsymbol{h}_{i,t}$ of an autoregressive recurrent neural network, $h$, with LSTM cells:

$$h_{i,t}=h(\boldsymbol{h}_{i,t-1},z_{i,t-1},x_{i,t},\theta) \tag{4.9}$$

Following Equation (4.9), at each time step $t$, the network $h$ takes as an input the features $x_{i,t}$, the target value at the previous step $z_{i,t-1}$ as well as its previous output $\boldsymbol{h}_{i,t-1}$ to produce $\boldsymbol{h}_{i,t}$ (and thus, allowing observation information to flow through time). The initial state $\boldsymbol{h}_{i,0}$ of the encoder as well as $z_i,0$ are initialized to zero.

The parameters of the likelihood $\ell(z|\theta)$ are given by a function $\theta(\boldsymbol{h}_{i,t},\Theta)$ of the network output ($\boldsymbol{h}_{i,t}$).One advantage is that this later distribution, which determines the *model noise*[4], can be chosen by the forecaster to match the statistical properties of the data at hand. For real value data, such distribution can be a Gaussian parameterized by $\theta=(\mu,\sigma)$ for instance.

Learning the parameters $\Theta$ of the model, comprising the parameters of the RNN $h(\cdot)$ as well as the parameters $\theta(\cdot)$, is done by maximizing the log-likelihood

$$\mathscr{L}=\sum_{i=1}^{N}\sum_{t=t_0}^{T_F}\log\ell\left(z_{i,t}|\theta\left(\boldsymbol{h}_{i,t}\right)\right). \tag{4.10}$$

The optimization can be, for example, carried out with stochastic gradient descent by computing the gradient of $\Theta$ since $\boldsymbol{h}_{i,t}$ is a deterministic function of the input.

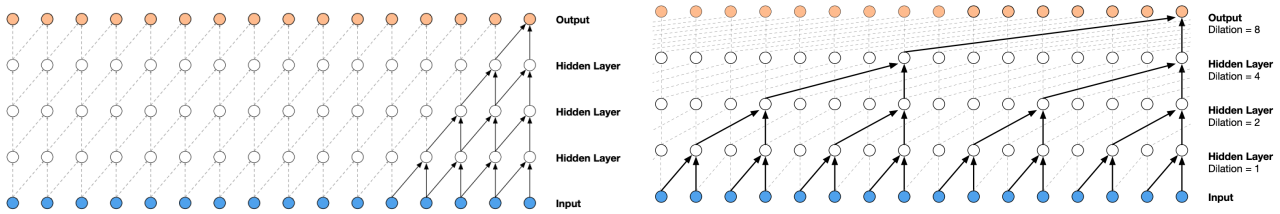
Given the parameter $\Theta$, DeepAR makes probabilistic forecasts in the form of $K\in\mathbb{N}_{>0}$ Monte Carlo samples $\tilde{\boldsymbol{z}}^k_{i,t_0:T_F}\sim Q_\Theta\left(\boldsymbol{z}_{i,t_0:T_F}|\boldsymbol{z}_{i,1:t_0-1},\boldsymbol{x}_{i,1:T_F}\right)$, $k=1,\ldots,K$. More specifically, the history of the time series, $z_{i,t}$ is first fed in for $t<t_0$. Then, for $t\geq t_0$, i.e. in the prediction range (right part of Figure (4.1)), a sample $\tilde{z}_{i,t}\sim\ell\left(\cdot|\theta\left(\boldsymbol{h}_{i,t},\Theta\right)\right)$ is drawn and fed back for the next point until the end of the prediction range $t=t_0+T_F$. The process is repeated to generate K samples from which we can derive any quantile.

## 2.4 Wavenet

Wavenet[5] is a generative deep neural network model based on Convolutional Neural Networks (CNNs) initially developed for speech analysis and audio waveform generation tasks. State of the art results (2016) have been obtained for text-to-speech problems. As such, this model seems less interesting at first glance and, therefore, won't be extensively covered even though it can also work for other time series related domains. The join probability of a time series (waveform) $\boldsymbol{z}_{1:T_i}=\{z_1,\ldots,z_T\}$ is factorized as a product of conditional probabilities

$$\boldsymbol{z}_{1:T}=\prod_{t=1}^{T}p(z_t|z_1,\ldots,z_{t-1},\boldsymbol{x}_{1:T_F}) \tag{4.11}$$

with each sample $z_t$ being conditioned on the samples of all the previous timesteps and additional, but optional, input variables $\boldsymbol{x}_{1:T_F}$ (e.g. the speaker identity). Wavenet models this conditional probability distribution by a stack of dilated causal convolutional layers as schematically picture in Figure (4.2). In a nutshell, causal convolutions make sure the model doesn't have a look ahead bias, i.e. that the output prediction $p(z_t|z_1,\ldots,z_{t-1},\boldsymbol{x})$ emitted at time $t$ indeed doesn't depend on future timesteps $z_{t+1},\ldots,z_T$. Dilated convolution, which skips input values with a certain step, are then introduced to increase the receptive field and mitigate the computation overhead introduced by the high number of layers required by causal convolutions.




(a) Causal convolutions.

(b) Dilated causal convolutions

Figure 4.2 – Schematic representation of a stack convolutional layers. Image sources: [5]

As a result, the higher-level dilated convolution layers can reach potentially far into the past to summarise lower levels and therefore acting as an alternative to long-term memory connections.

Multi-step predictions are done in a sequential manner, by feeding back the one step prediction sample into the network to predict the next sample. Interestingly though, the model outputs a categorical distribution over the next value thanks to a flexible Softmax layer, with the promise that it better models arbitrary distributions ( $p(z_t|z_1,\dots,z_{t-1},\boldsymbol{x})$ since it doesn't make any assumptions about their shapes. The model is optimized to maximize the log-likelihood of the data with regards to the parameters.

## 2.5 DeepState

The main idea behind DeepState[62] is to parameterize a per-time-series linear State Space Model[46] (SSM) with a RNN whose weights are learned jointly from all the time series and covariates. On the one hand, traditional SSMs are interpreteable models, which don't require very long history. But since they are fit on individual time series, model and covariate selection turn to either a compute-intensive or prove to require human labor intensive work and they obviously can't infer shared patterns across *similar* time series. On the other hand, deep learning features a set of black box models which are able to extract high order features, identify complex patterns by making fewer structural assumptions at the cost of more training data (overfitting tendency). DeepState is an interesting proposition to include in our study since it strives to take the best of both worlds and to counterbalance each of their flaws.

State space models model the *temporal structure* by encoding time series components (e.g. level, trend and seasonality patterns), into a latent space $\boldsymbol{l}_t \in \mathbb{R}^D$. It is described by a *state-transition* equation defining the stochastic transition dynamics $p(\boldsymbol{l}_t|\boldsymbol{l}_{t-1})$ by which the latent state evolves over time as well as an observation model, specifying the conditional probability $(z_t|\boldsymbol{l}_t)$, which describes how the observations are generated from the latent space $\boldsymbol{l}_t$[62]. For instance, a linear innovation SSM yields the following transition dynamics and observation model (we omit the time series index $i$):

$$\begin{aligned} &\boldsymbol{l}_t = \boldsymbol{F}_t\boldsymbol{l}_{t-1} + \boldsymbol{g}_t\varepsilon_t && \text{with } \boldsymbol{l}_0 \sim \mathcal{N}(\mu_0,\Sigma_0), \text{ and } \varepsilon_t \sim \mathcal{N}(0,1) \\ &z_t = y_t + \sigma_t\epsilon_t, \quad y_t = \boldsymbol{a}_t^\top\boldsymbol{l}_{t-1} + b_t, && \text{with } \epsilon_t \sim \mathcal{N}(0,1) \\ &\boldsymbol{a}_t \in \mathbb{R}^L, \sigma_t \in \mathbb{R}_{>0}, b_t \in \mathbb{R} \end{aligned} \tag{4.12}$$

where $\Theta_t = \left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \boldsymbol{F}_t, \boldsymbol{g}_t, \boldsymbol{a}_t, b_t, \sigma_t\right), \forall t > 0$ are parameters of the model. $F_t$ is actually a *deterministic* transition matrix and $\boldsymbol{g}_t\epsilon_t$ the *random innovation* of strength $\boldsymbol{g}_t$.

In a nutshell, DeepState learns a global mapping $\Psi\left(\boldsymbol{x}_{i,1:t}, \Phi\right)$ from the covariate vectors $\boldsymbol{x}_{i,1:T_i}$ associated with $z_{i,1:T_i}$ to the time varying parameters $\Theta_{i,1:T}$ of the linear state space models of each time series $i = 1,\dots,N$). The mapping is parameterized by a RNN with LSTM layers with parameters $\Phi$ which generates the hidden states:

$$\boldsymbol{h}_{i,t} = h\left(\boldsymbol{h}_{i,t-1}, \boldsymbol{x}_{i,t}, \Phi\right) \tag{4.13}$$

In this way, the output values of the last LSTM layer is mapped to the parameters $\Theta_{i,1:T}$ of the state space model through a set of element-wise transformation constraining the parameters to valid ranges. The model is trained so as to maximize the probability of observing the data in the training range (log-likelihood):

$$\Phi^\star = \text{argmax}_\Phi \mathscr{L}(\Phi) \quad \mathscr{L}(\Phi) = \sum_{i=1}^{N} \log p\left(\boldsymbol{z}_{i,1:T_i} | \boldsymbol{x}_{i,1:T_i}, \Phi\right) \tag{4.14}$$

where each term of Equation (4.14) measures the compatibility between the SSM parameters $\Theta_{1:T_i}$ produced by the RNN when given input $\boldsymbol{x}_{i,1:T_i}$[62].

Finally, the prediction trajectory $z_{i,1:T_i}$ is distributed according to

$$p\left(\boldsymbol{z}_{i,1:T_i} | \boldsymbol{x}_{i,1:T_i}, \Phi\right) = p_{SS}\left(\boldsymbol{z}_{i,1:T_i} | \Theta_{i,1:T_i}\right), \quad i = 1, \ldots, N \tag{4.15}$$

where $p_{SS}$ is the marginal likelihood under the linear state space model of parameters $\Theta_{i,1:T}$[62].

## 2.6 MQ-RNN and MQ-CNN

So far, all the models described relied on a recursive prediction scheme to generate multi-step forecasts, which, as already mentioned, lead to error accumulation. The next and final model we'll describe sets apart by embracing a direct strategy to predict the multivariate target $(z_{t_0}, \ldots, z_{T_F})$ given $\boldsymbol{z}_{1:t_0-1}$. Proclaimed advantages of the multi-horizon approach is a bias reduction, increased stability and robustness improvements to model mis-specifications[63, 6].

In this way, Multi-Quantile RNN[6] (MQ-RNN or variant MQ-CNN) improves upon seq2seq networks by leveraging Quantile Regression (QR) which learns to predict the conditional quantiles $z_{t+h}^{(q)} | \boldsymbol{z}_{1:t}$ of the target distribution $\mathbb{P}\left(z_{t+h} \leq z_{t+h}^{(q)} | \boldsymbol{z}_{1:t}\right) = q$ for each $h = 1, \ldots, \tau$. By doing so, the model doesn't make any distributional assumptions (e.g. Gaussian residuals) which makes it more robust. Had the model be presented in the GefCom2014[43], it would have scored first place on the price forecasting track. For all the above reasons, the model is a prime candidate to include in our empirical study.

MQRNN (MQCNN) handles cross-learning between the distinct time series $\left\{\boldsymbol{z}_{i,1:T_i}\right\}_{i=1}^{N}$ by means of a single RNN(CNN) where each time series being considered as one sample. This has the further benefit of allowing cold-start forecasting when the history of some series might be limited. Consequently, we drop the index $i$ to simplify the notation.

**Training** QR models are trained to minimize the so-called Quantile Loss (QL):

$$L_q(z, \hat{z}) = q\,\texttt{max}(z - \hat{z}, 0) + (1 - q)\,\texttt{max}(\hat{z} - z, 0) \tag{4.16}$$

Denoting by $\tau$ the number of prediction horizons and $Q$ the number of quantiles of interest, the parametric model $h(z_{1:t}, x, \theta)$ (e.g. an RNN or CNN) outputs a single matrix $\hat{Z} = \left[\hat{z}_{t+h}^{(q)}\right]_{h,q}$ of dimensions $\tau \times Q$. The parameter of $h$ are trained to minimize the sum of individual quantile loss and is given by:

$$\mathscr{L} = \sum_t \sum_q \sum_h L_q\left(z_{t+h}, \hat{z}_{t+h}^{(q)}\right) \tag{4.17}$$

where $t$ iterates over all forecast creation times[6].

**Architecture** Figure (4.3) shows the neural network architecture in the case of MQRNN. It uses a vanilla RNN-based encoder which encodes hidden states $\boldsymbol{h}_t$ with LSTMs cells. Interestingly, this encoder can even be replaced by a CNN-based encoder like Wavenet which leads to the MQ-CNN variant. However, the encoder is not followed by a standard recursive LSTM based decoder as usual but by two custom MLP branches.

In this way, a first (*global*) MLP, $m_G$, is responsible for summarizing the information contained in the encoder output $\boldsymbol{h}_t$ and future features $\boldsymbol{x}_{t:t+\tau}$. This MLP then produces a series of horizon-specific contexts $\boldsymbol{c}_{t+h}$ as well as a general (horizon-agnostic[6]) context $\boldsymbol{c}_a$ which captures common information:

$$(\boldsymbol{c}_{t+1}, \cdots, \boldsymbol{c}_{t+\tau}, \boldsymbol{c}_a) = m_G\left(\boldsymbol{h}_t, \boldsymbol{x}_{t:t+\tau}^{(f)}\right) \tag{4.18}$$

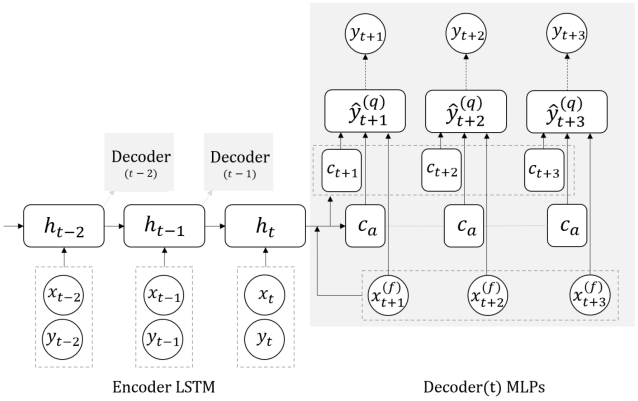



Figure 4.3 – MQRNN multi-step forecast architecture. $y_t$ corresponds to $z_t$ according to our notation. *Source:*[6].

The second MLP, $m_L$ is said to be *local* since it applies only to each specific horizon. Its parameters are, like the first MLP, time-invariant. Indeed, while a series of such decoders is placed at each recurrent layer of the encoder, their parameters are actually shared across all horizons $h \in \{1,\dots,\tau\}$. Each produces the configured quantiles from the information set $\left(\boldsymbol{c}_{t+k}, \boldsymbol{c}_a, x^{(f)}_{t+k}\right)$:

$$\left(\hat{y}^{(q_1)}_{t+h}, \cdots, \hat{y}^{(q_Q)}_{t+h}\right) = m_L\left(\boldsymbol{c}_{t+h}, \boldsymbol{c}_a, \boldsymbol{x}^{(f)}_{t+h}\right) \tag{4.19}$$

# 5 | Data Analysis

## 1 Market data visualization

### 1.1 CAL visualizations

A first visualization of the available data (Figure (2.7)) is carried out with *plotly* in order to generate interactive html/js graphs that can be more easily incorporated into a future dashboard. In Figure (5.1), all the CALs are plotted on separate graphs with common axis.

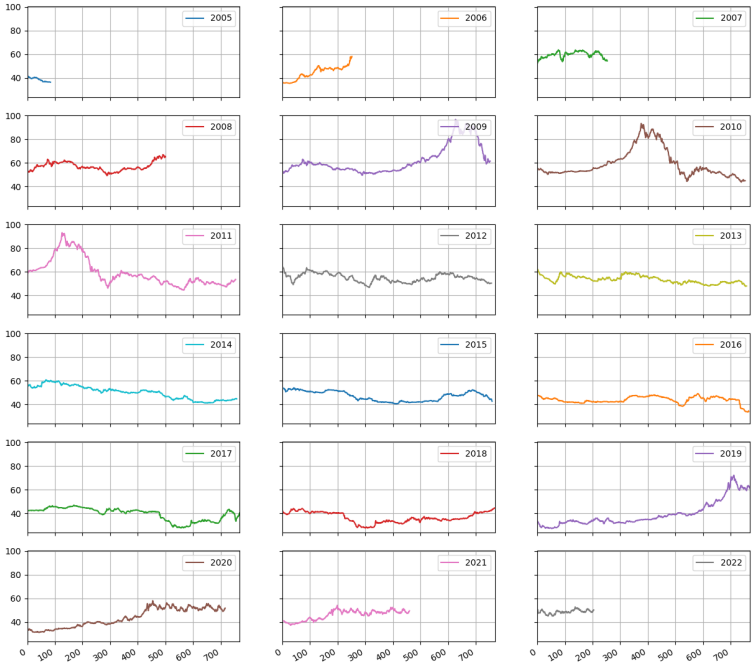



Figure 5.1 – Grid plot of the available year-ahead data. [x]: open days, [y]: €/MW

We directly notice that the full quotation is missing for old CALs like 2005, 2006, 2007 and 2008. In a similar way, CALs 2020, 2021 and 2022 deal with delivery in the future at time of writing and thus can't be extended over 3 years of quotation. The number of open days also varies from one CAL to another. Furthermore, trends in overlapping and concurrent[1] CALs seem to be highly correlated. A steep increase during the financial crisis of 2008 (quotation time) is also clearly noticeable on CALs 2009 to 2011.

Figure (5.2) shows that the price distribution can change a lot from a CAL to another. Some years, like 2012, seem to follow a symmetric distribution. This can be checked by noticing that the median is located in the middle of the Interquartile Range (IQR) and that the whiskers are symmetric too. However, most of the CALs seem to follow a skewed distribution. The median is not centered in the interquartile range, and many outliers, outside of the $1.5 \times \texttt{IQR}$ can sometimes be noticeable. The reduced variance for CAL 2005, CAL 2006, CAL 2007, CAL 2021 and CAL 2022 is a consequence of less data points available.

[1] understand CAL being quoted at the same time

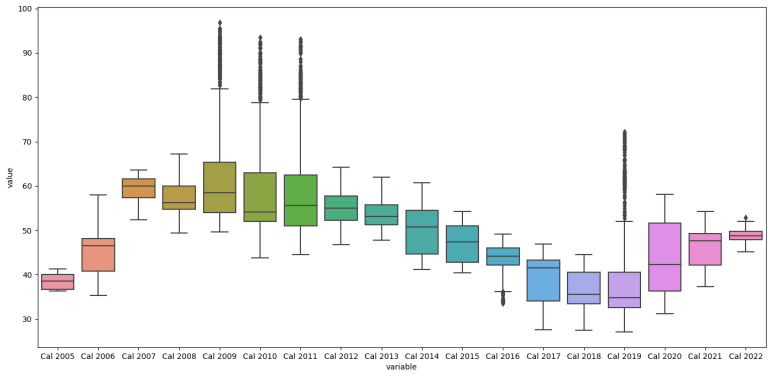



Figure 5.2 – Boxplots of year-ahead CAL.

## 1.2 Market interruptions

From Figure (5.3), we notice that, between closing and opening days, some gaps arise. For example, we can observe that a gap of 10% can arise between a closing and an opening of the market.

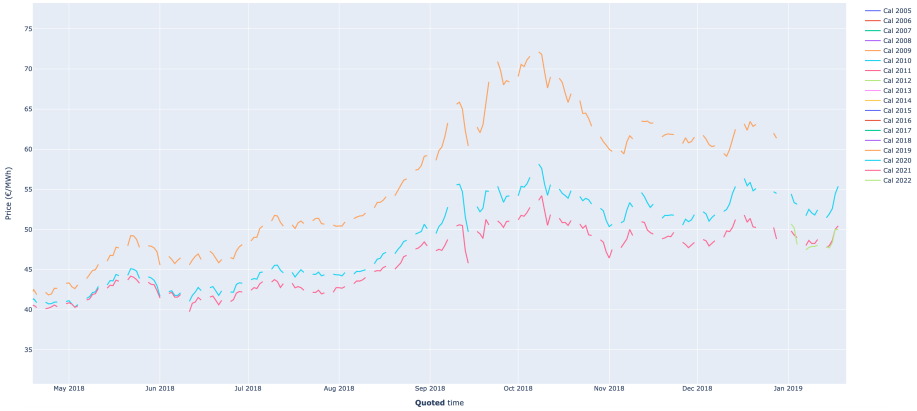



Figure 5.3 – Prices jump between closing and next opening of the YAM market.

Finally it should also be mentioned that the beginning of the quotation period for a given CAL $y$, doesn't always start in the first open days of January of year $y-3$. Some CALs have indeed their first open days at the end of December of $y-4$ in the provided data.

## 1.3 Smoothing and filling missing entries

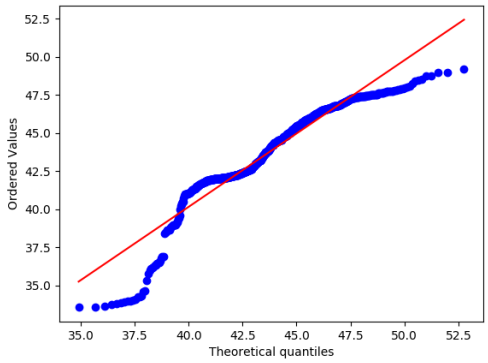
The year-ahead market data turned out not to require any form of data cleaning. When filling missing entries (closed market) is required, for forecasting for instance, it is done with a simple *forward fill* by default to preserve as much as possible the original data. Various smoothing techniques can also be applied. The objective behind the idea of smoothing the time serie is three folds. First, it helps better visualizing the underlying trend. Second, and most important, we hope that smoothing would get us a fair approximation of the noise-filtered series. Finally, the smoothed version of series can be used as a technical indicator (feature) for the original time series. For instance, economists leverage several smoothing filters with different lengths (50-200 most of the time) and consider their crossing points as indicator that a reversal has occurred and thus to generate buy/sell signals. We refer to the Appendix 5.1 for a brief explanation and discussion on the most common filters that are moving average, median average, LOESS and EWMA. Figure (8.8) also gives a visual representation of CAL 2016 smoothed according to the mentioned filters.

# 2 Statistical market data analysis

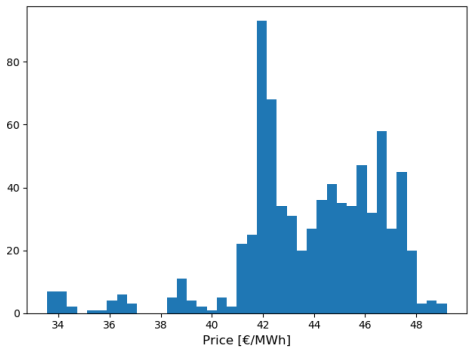
## 2.1 Normality assumptions

Meeting the normality assumptions has generally a positive impact on regression models and estimators. However, most of them can also handle fairly non-normal data as long as the residuals are fairly normal.

A Quantile-Quantile (Q–Q) plots is a graphical tool that can be used to compare the properties like location, scale and skewness of two distributions. Figure (5.4a) graphically compares the sample quantiles of the price data with the theoretical quantiles of a conventional normal distribution. The tails of the distributions do not seem to align very well, therefore, a normality assumption doesn't seem to hold.




(a) Normality check with QQ plot.




(b) Histogram of prices.

Figure 5.4 – Normality analysis on CAL 2016.

To confirm this statement in a more formal way, a normality test, whose null hypothesis $H_0$ corresponds to the samples having a normal distribution, is made. Table 5.1 shows $H_0$ is indeed rejected since $p$ is below the significance level $\alpha = 0.005$. The *kurtosis* parameter is a measure of the combined weight of the tails relative to the rest of the distribution. In other words, it measures the tail-heaviness of the sample distribution with zero corresponding to the value obtained for a normal distribution. On the other hand, *skewness* measures the lack of symmetry of the distribution. A negative skewness would correspond to a longer left tail while a positive skewness is interpreted as the reverse. Thus, Table 5.1 tells us that the price distribution is indeed heavy-tailed (as stated in [39])and asymmetric.

| Statistics | p | Reject H0 ? | Kurtosis | Skewness |
|---|---|---|---|---|
| 131.886 | 0.000 | ✓ | 1.756 | -1.037 |

Table 5.1 – Normality test on cal 2016 prices

## 2.2 Time dependencies and patterns

### 2.2.1 Lag

Figure (5.5) plots the prices values against a shifted version of the time series. Finding lags such that a line pattern emerges would prove there is a time dependency between a value and its past observations. Since, it is the case, we can safely say prices are not random. To get more information on the *correlation* between different lagged versions of the time series, we then test its autocorrelation.

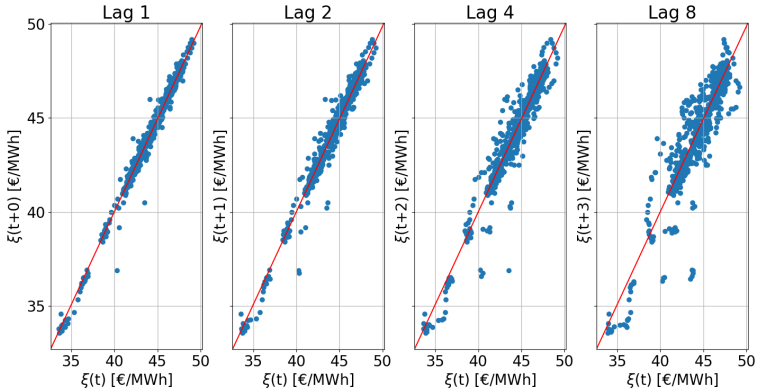



Figure 5.5 – Lag plots (CAL 2016).

### 2.2.2 Autocorrelation

The *autocorrelation* generalizes the notion of covariance by taking the Pearson correlation between values of the process at different times. It is nothing else than the correlation of the series with a delayed copy of itself as a function of this lag $k$:

$$\rho_{\xi\xi}(k) = \frac{\mathbb{E}\left[(\xi(t)-\mu)(\xi(t+k)-\mu)\right]}{\sigma^2} \tag{5.1}$$

It is a common practice to normalize the autocorrelation function, so that its value must fall in the range $[-1,1]$, with 1 indicating perfect correlation and -1 indicating perfect anti-correlation. Computing the autocorrelation of each individual CAL does not highlight any relevant *periodicity* pattern. Figure (5.6) exposes the results found for CAL 2016. Figure (5.6a) shows that the future values have high correlation with previous values but that there are no obvious underlying patterns in the time series. The Pearson coefficient is positive and decreases linearly up to a lag close to 30. Thereafter, the autocorrelation is not significant (Figure (5.6b)) and crosses the zero axis near 80 to get positive again, but not significantly either when applying a lag of around 230. Most of the other CALs have instead a "U" shape that crosses the 0 line only twice.

The partial autocorrelation of lag $k$ denoted $\alpha(k)$ is the autocorrelation between $\xi_t$ and $\xi_{t+k}$ with the linear dependence of $\xi_t$ on $\xi_{t+1}$ through $\xi_{t+k-1}$ removed. In other words, it is the autocorrelation between $\xi_t$ and $\xi_{t+k}$ that is not accounted for by lags 1 through k-1 inclusive[64] with $\alpha(1) = \rho(\xi_t, \xi_{t+1})$. Figure (5.6c) clearly highlights that two consecutive values are highly correlated but that this correlation collapses suddenly past a lag of 1. The partial autocorrelation for a lag of one amounts to 0.98. Interpretation of this finding will be discussed in Subsection 2.3.

### 2.2.3 Seasonality, trend and residuals decompositions

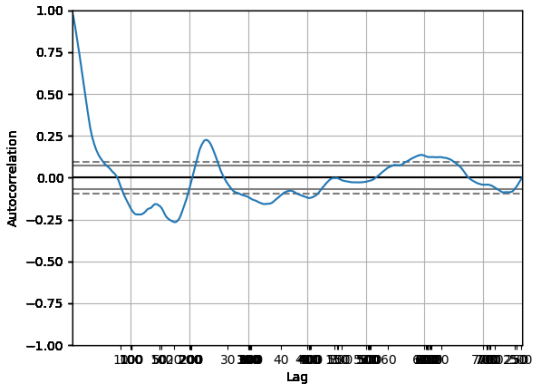
A time series can be thought as the combination of four components:

- $Level$: the mean of the series.
- $Trend$ (optional): the increasing or decreasing value in the series.
- $Seasonality$ (optional): the repeating (short-term or medium-term) cycle in the series.
- $Noise$: the random variation in the series.

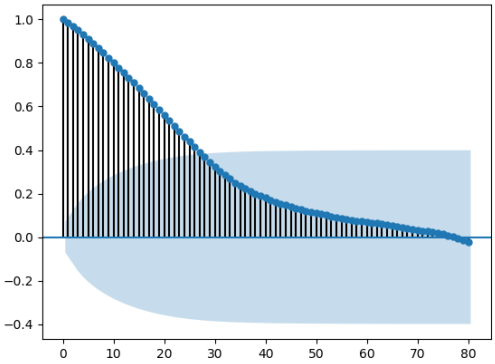
either via an additive model $\xi(t) = level(t) + trend(t) + season(t) + noise(t)$ or a multiplicative model $\xi(t) = level(t) * trend(t) * season(t) * noise(t)$.

Figure (5.7a) is the result of the `seasonal_decompose()` function of the *statsmodel* Python package for an additive model. The trend is extracted using an automatically tuned moving average filter[2] so that the esti-

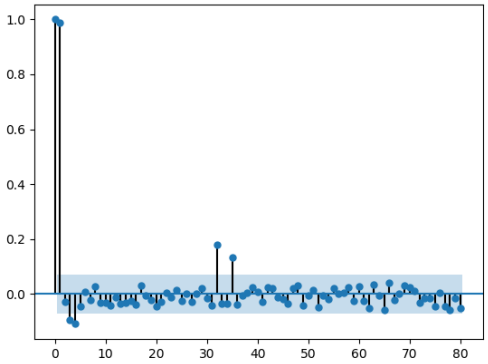
[2] Even length are handled by weighting the two extreme lags by $\frac{1}{2}$




(a) Autocorrelation for all possible time shifts.



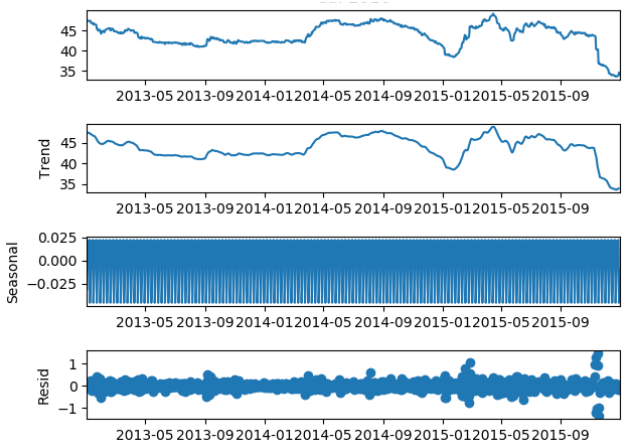
(b) Autocorrelation for up to 80 time shifts.



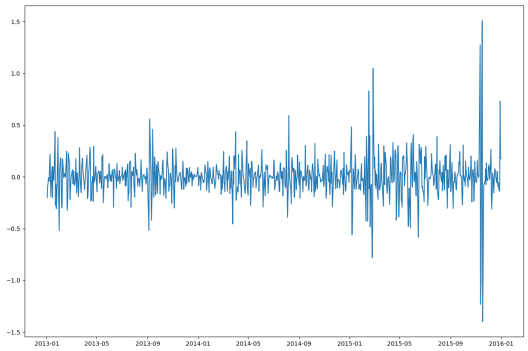
(c) Partial autocorrelation for up to 80 time shifts.

Figure 5.6 – Autocorrelation analysis of CAL 2016.

mation of the trend computed from from lag $-a$ to $a$ is given by $\hat{m}(t) = \sum_{k=-a}^{a} \left(\frac{1}{1+2a}\right) \xi(t+k)$. The seasonal effect is then given by $\hat{s}(t) = \xi(t) - \hat{m}(t)$ while the random error term is the residual $\hat{e}(t) = \xi(t) - \hat{m}(t) - \hat{s}(t)$. As a result, the detrended line $\hat{s}(t)$ also includes the random error term[65].




(a) Additive seasonality decomposition.



(b) Detrended time series.

Figure 5.7 – Time series decomposition (CAL 2016)

Figure (5.7b) shows that the time serie has no seasonal component but results should be taken with a little caution since the decomposition is still an approximation. It is widely recognized that seasonality is often present in business related time series[60]. However, considering that the purchased electricity quantity is to be delivered at a constant rate for a given year, the absence of seasonality is not surprising at all. It can

be easily understood that there are no reasons that holidays or week-ends have any influence on the price of the base load supply. Plotting a bar graph of the average year-ahead prices per weekday would not show any useful pattern which is in complete opposition to what can be observed for the DAM (see Appendix Figure (8.9) for a discussion). One way to exploit this decomposition is to forecast separately the seasonally adjusted time series using any model and the seasonal component using a seasonal naive method of the form $\hat{y}(t+h|t) = y(t+h-T)$ where $T$ designates the period of the seasonality and $h$ the forecast horizon[46]. However, we decided not to proceed in this way due to low contribution of the seasonal component which might as well suffer from approximations.

## 2.3 Stationarity and differentiation

A stationary process has the property that its statistical properties, like mean and variance, are constant (independent of the time index) over time. Moreover, its *autocorrelation* structure should be constant over time nor should it have a periodic fluctuation. We say the later is said to have no *trend* fluctuation and no *seasonal* component. Strict stationarity, in the mathematical sense, imposes that the joint distribution of observations is invariant to time shifts.

Stationarity can be checked visually with plots (see Appendix **??**) or more objectively by running a statistical test. The Augmented Dickey-Fuller (ADF) test[66, 67], with the null hypothesis,

$H_0$ : *The time series possesses a unit root[67] and is non-stationary*

is the most commonly used one[66]. For instance, if the P-Value in the ADF test is less than the significance level (0.05), we reject the null hypothesis. In Table 5.2, we notice that the time series do not satisfy the stationarity property since, for ADF, the p-value is greater than all the significance levels.

| **Test** | **ADF** |
|---|---|
| Statistic | -2.342914 |
| p-value | 0.158500 |
| **Critical values** | |
| 1% | −3.439 |
| 5% | −2.865 |
| 10% | −2.569 |

Table 5.2 – Stationarity statistical tests

Some autoregressive models work better if the observations are not correlated against each others[46]. Stationarizing removes any persistent autocorrelation and makes the predictors (lags of the series, see Equation (2.1)) in the forecasting models nearly independent. A convenient method to stationarize the data is by differencing the series until it becomes approximately stationary (see Figure (5.9b)). The first difference CAL is obtained by replacing an observation $\xi(t)$ by its difference with the previous timestamp, i.e. $\xi'(t) = \xi(t) - \xi(t-1)$.

Figure (5.9b) shows the differentiated prices of CAL 2016, the mean and variance of which are respectively -0.01782 €/MWh and 0.09913 (€/MWh)$^2$. The resulting time series has then a close to non-existent autocorrelation with itself which is unsurprising considering the first difference data looks graphically almost like white noise[3]. This is in line with Figure (5.6c) which showed that the maximum correlation was achieved with the next value before collapsing suddenly with a $lag = 2$. We can thus expect that the first difference values are close to being independent to one another. If the later is true, the element-wise (partial) correlation of the first difference data would be close to zero for all $lag \neq 0$. Figure (5.10) shows it is actually the

[3] A white noise is a discrete signal whose samples form a sequence of serially uncorrelated random variables with zero mean and finite variance.

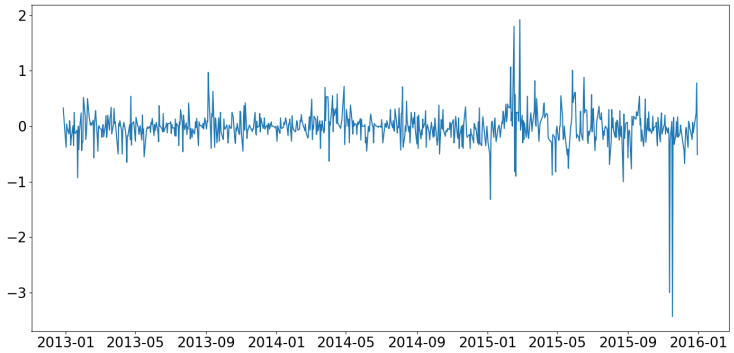
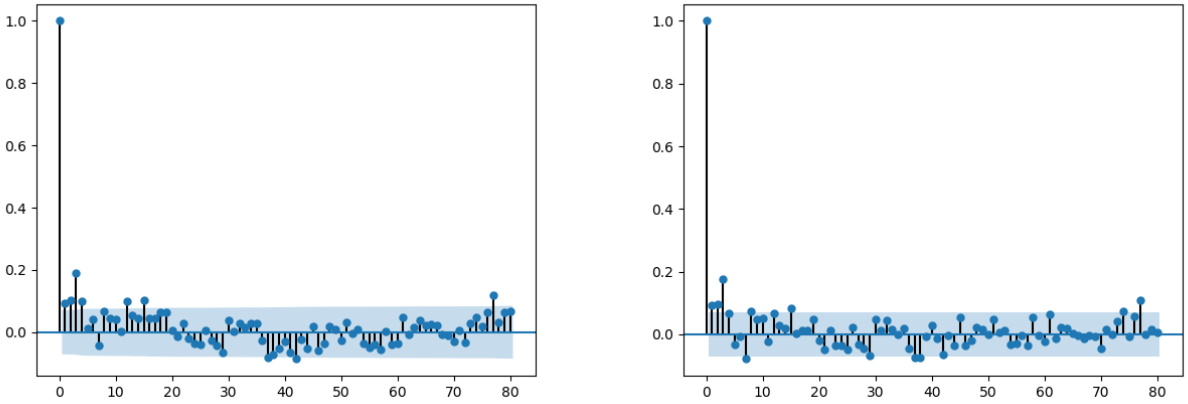



Figure 5.8 – First difference of CAL 2016




(a) Autocorrelation.

(b) Partial autocorrelation.

Figure 5.9 – Correlation of CAL 2016 first difference prices for up to 80 time shifts.

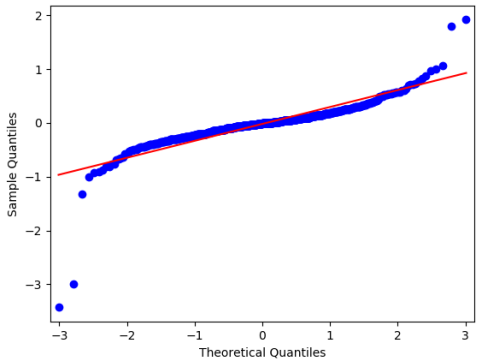
case. By running again an ADF test, we get an ADF statistic of -5.123 with a p-value of 0 and thus the H0 hypothesis is rejected: the resulting process is stationary.

Plotting the histogram of the first difference reveals its distribution is bell-shaped and centered on 0. However, from Figure (5.10a), we notice it does not satisfy the normality assumption as the curve of the distribution is too narrow and has exceptional outliers[4]. This is inline with the high curtosis found for the statistical test. A distribution with a kurtosis greater than three is qualified as *leptokurtic*. Such distributions are known to produce less reliable and less safe value at risk distributions since the worst scenarios, derived from the long left tail, are located further from the mean than the one estimated with a normal assumption. As a result, using a normal hypothesis would be too optimistic with regards to the worst case scenario but at least would be pessimistic with regards to the first difference price variance in general.

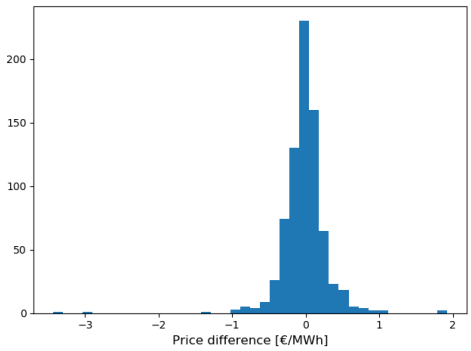
| Statistics | p | Reject H0 ? | Kurtosis | Skewness |
|---|---|---|---|---|
| 499.986 | 0.000 | ✓ | 31.477 | -2.198 |

Table 5.3 – Normality test on first difference cal 2016 prices

[4]Removing the 27 highest peaks among 766 data points, which corresponds to 3.5% of the data makes the normality assumption correct though.




(a) Normality check with QQ plot




(b) Histogram of differentiated prices

Figure 5.10 – Normality analysis on first differentiated CAL 2016

## 2.4 Evaluation of forecastability

Figure (5.5) highlighted a decaying time dependency in the data while Figure (5.6c) showed peak correlation between two consecutive values. The time series is not stationary and the analysis on the first differenciated data highlighted a mean close to zero, with almost no autocorrelation. Unfortunately, all these properties are very similar to what we would obtain for a random walk process.

As already introduced in Subsection 2.1, an autoregressive (AR(p)) model is a representation of a type of random process which has the form:

$$X_t = c + \sum_{i=1}^{p} \alpha_i X_{t-i} + \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{5.2}$$

where $\alpha_1, \ldots, \alpha_p$ are parameters of the model and $\epsilon_t$ is an *iid* white noise with a normal distribution $\mathcal{N}(0, \sigma^2)$. From the computed partial autocorrelation (Figure (5.6c)), we know that $p = 1$ in our case and the model can be simplified as:

$$X_t = \alpha X_{t-1} + \epsilon_t,\ \epsilon \sim \mathcal{N}(0, \sigma^2). \tag{5.3}$$

When $|\alpha| \leq 1$ , the random process is non-stationary, it is a pure random walk when $\alpha = 1$, otherwise it is explosive. Like our time series, random walks are non-stationary. Indeed, by resolving the recursive definition for an AR(1) model, i.e. a pure random walk process, with $\alpha = 1$, we get:

$$\begin{aligned} X_t &= X_{t-1} + \epsilon_t && (5.4)\\ &= X_{t-2} + \epsilon_{t-1} + \epsilon_t && (5.5)\\ &\vdots && (5.6)\\ &= X_0 + \sum_{i=0}^{t-1} \epsilon_{t-i} && (5.7) \end{aligned}$$

Then, by considering the expectation of the random variable $X_t$, we get:

$$\mathbb{E}[X_t] = \mathbb{E}[X_0] + \sum_{i=0}^{t-1} \mathbb{E}[\epsilon_{t-1}] = \mathbb{E}[X_0] \tag{5.8}$$

since the second term can be dropped due to the iid sampling with a mean centered in 0. Thus, if $\mathbb{E}[X_0] = 0$, we naturally get $\mathbb{E}[X_t] = 0$. The consequence of which is that the process has a constant mean over time. However, by considering f.e. $\mathbb{E}[X_0] = 0$, the variance is not. Indeed, we have:

$$Var[X_t] = \sum_{i=0}^{t-1} Var[\epsilon_{t-i}] = t\sigma^2 = f(t) \tag{5.9}$$

where the covariance term can be dropped since the sampling is done iid. Equation (5.9) clearly depends on time and the variance diverges to infinity with $t$. Hence the process is non-stationary.

We already showed the prices from CAL 2016 shared many common properties with the one inherent to random walk processes. **??** illustrates an artificially generated random walk process alongside all the previously introduced plots for this random instance. Visually, all the analysis plots show a striking resemblance to the one found for CAL 2016. The study could have been likewise conducted on other CAL but similar results would have been obtained. The same summary plots for 2013, 2015, 2017 and 2019 are given in Appendix 5.3 for the interested reader.

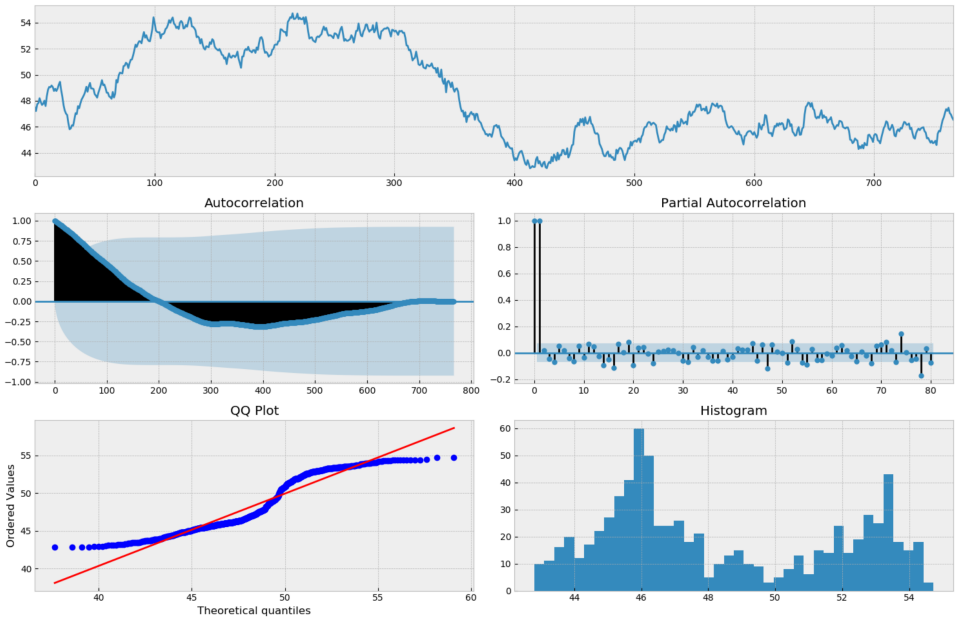



Figure 5.11 – Artificially generated random walk (AR(1) process). Noise sampled according to $\epsilon \sim \mathcal{N}(-0.0178, 0.0991)$ with $\xi_0$, $\mu$ and $\sigma^2$ set to match the one of CAL2016.

### 2.5 Conclusion on statistical data analysis

Time series with regular patterns are obviously known to be easier to forecast. Unfortunately, we weren't able to find any in the data. Forecasting the future price distribution is expected to be very challenging, since, due to the highlighted AR(1) property of the data, a naive predictor of the form $\hat{x}_{t+k} = x_t$, $\forall k \in \{1, \ldots, |T_F - t|\}$ would be among, if not, the best estimator for a random walk like process. Let's keep in mind this observation holds for the time series as a whole.

## 3 Correlation and causation between time series

Visualizing the interactions between concurrently quoted CAL in the YAM electricity prices dataset can provide relevant insights on the dynamic of the market. Furthermore, in lights of Subsection 3.3, several exogenous factors influence electricity prices. Professional software companies like Aleasoft[33] and N-Side[34] also leverage thousands of variables to improve their long-term forecast accuracy on short-term delivery electricity markets. Thus, we first expose very briefly a set of tools from the data mining community that can be used to evaluate the dependencies between time series and their mutual contribution to improve forecastability. We then analyze the similarities and causations relations between concurrently quoted cal and a few exogenous variables with the hope to counterbalance the conclusion of the previous section.

## 3.1 Background

### 3.1.1 Global correlation

We consider *similarity* as the measure that establishes an absolute value of resemblance between two vectors. The (normalized) cross-correlation function, or *Pearson correlation coefficient*, is among the most widely used measure of similarity due to its ease of interpretation and yet practical effectiveness[68]. The Pearson coefficient $r_{XY}$ between two variables X and Y is defined as the covariance of the two variables divided by the product of the standard deviation of each data sample:

$$r_{X,Y} = \frac{cov(X,Y)}{\sigma(X).\sigma(Y)} = \frac{\mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])]}{\sigma(X).\sigma(Y)} \in [-1,1] \tag{5.10}$$

where the denominator adjusts the scales of the variables to have equal units.

The covariance, like the correlation, is a symmetric measure of linear association between variables[68]. The use of mean and standard deviation in the calculation suggests the need for the two data samples to have a Gaussian-like distribution. As a consequence, the Pearson coefficient measures the strength of the linear relationship between normally distributed variables. A value for $r_{XY} = +1$ would correspond to a perfect positive linear correlation, 0 to no linear correlation and -1 to a negative correlation. In case linearity or normality assumptions do not hold, the nonparametric *Spearman rank correlation* $\rho_{XY}$ can be used to measure monotonic relationships (linear or not) instead. The later is simply defined as the pearson correlation coefficient between rank variables[69]. In our case, the lesser assumptions, the better so the Spearman correlation should be preferred to assess how much the time series are related to one another.

### 3.1.2 Global causality

The above mentioned coefficients don't provide yet any insights on dynamics such as leader-follower relationships[5]. Identifying the direction of such dynamics can be achieved with *Time Lagged Cross Correlation* (TLCC)[70]. It simply consists in computing repeatedly a correlation coefficient between one of the time series and an incrementally shifted version of the other one where both lag directions are explored. The lag value for which the peak correlation is reached indicates the lag which corresponds to the best synchronization of the two time series.

However, in economics, a more complex statistical concept known as the *Granger Causality*, first proposed in [71], is traditionally preferred[72]. Let $Y_K$ be a lagged-k version of $Y$. In a nutshell, $X$ is said to *Granger causes* $Y$ if the optimal linear predictor $Y_k$, based on the information set comprising X, has a smaller variance than the one on $Y_k$ alone. Through a series of statistical tests for lag $k$, it is then the practice to show that $X$ provides statistically significant information about the future of $Y$.

### 3.1.3 Assessment limitations

The above mentioned techniques only assess global relationships on the common time range of several time series. However, the exposed techniques do not truly capture local evolution of the causality relationships which might evolve over time. An easy extension would be to perform the same analysis on a rolling window basis but would require a lot of human labor for several time series. We instead simply assesses global relationships. *Heteroscedasticity*[6] and outliers are additional factors that can negatively impact the quality of the study[70].

[5] Situations in which a leader series initiates a response which is repeated by a follower

[6] *Heteroscedasticity* is the property by which the variance of error in a time series is heterogenous across the time range.

### 3.2 Concurrently quoted CAL

Zooming on Figure (5.3) shows that the price movements of concurrently quoted CAL seems to be highly correlated. To confirm this intuition, Table 5.4 summarizes the correlation coefficients computed on the raw time series as well as the time series replaced by the one step percent change. The later simply replaces each value y(t) according to:

$$y(t) = \frac{y(t-1) - y(t)}{y(t)}. \tag{5.11}$$

We notice highly positively correlation. The correlation on the percent change is however reduced and the pairwise ordering in terms of similarity is not the same. One explanation is that ordinal values contain level information that are lost by introducing the transformation.

| Similarity Measure | Percent change data | cal2015 - cal2016 | cal2016 - cal2017 | cal2015 - cal2017 |
|---|---|---|---|---|
| Pearson correlation | no | 0.892 | 0.968 | 0.905 |
| Spearman correlation | no | 0.850 | 0.954 | 0.876 |
| Pearson correlation | yes | 0.634 | 0.595 | 0.499 |
| Spearman correlation | yes | 0.600 | 0.538 | 0.490 |

Table 5.4 – Similarity measure between concurrently quoted cal.

### 3.3 Exogenous data

Exogenous variables susceptible to improve forecastability are numerous. Most of the time, calendar data is used but this is irrelevant in our situation. Changes in regulation are not really predictable. The information about planned maintenance fetched from the ENTSO-E REST API does not bring relevant information either due to to short time horizon. Year-ahead commodity prices from ICE are notoriously difficult to get hands on since they are very expensive. Instead, we proactively collect information from freely accessible databases and expose our findings for a set of plausible exogenous variables. We simply redirect to the provided data sources for more information on them.

**Commodity prices** Carbon pricing is an instrument that captures the external costs of greenhouse gas (GHG) emissions. In lights of the discussion in Subsection 3.1 and Subsection 3.3, we expect carbon prices to influence electricity prices more and more in the future due to the increasing share of renewable energy production units and the rapid increase of carbon taxes (starting from 2018). The dataset considered is the *Futures EU allowance*[73] contracts from ICE: a discussion of which and a scaled visual comparison with YAM prices is exposed in Appendix 5.4.1. Besides, several *oil prices* (weighted and taxes removed)[74] are available on the Europa platform and are included in the analysis. The data frequency is weekly. No freely accessible information sources have been found for other energy related commodities.

**Financial Market indexes** A market index aims at measuring the performance of the considered market and its underlying stocks as a whole. Only the *S&P 350 Europe*[75] and the *BEL20*[76] markets are considered here. The European *consumer price index*[77], used to measure inflation, is only available at a monthly frequency and is thus not considered.

**Energy indexes** More specialized indexes for the energy sector exist. The *STOXX EU Oil & gas (SXEP) index* [78] benchmarks integrated oil and gas companies engaged in the exploration, drilling, production, refining, distribution and retail sales of oil and gas products. Finally, we also include the *Euronext Rogers*

*International Energy Commodity Index* [79] which benchmarks usual future contracts for energy commodities like crude oil, gas oil, brent oil and natural gas to list a few. The origin of the underlying securities come from ICE and NYMEX.

Figure (8.14), available in Subsubsection 5.4.2 displays all the logged transformed, weekly aggregated and scaled in [0, 1] data. Figure (5.12) displays the correlation coefficients obtained after scaling all the augmented dataset in [0, 1] and aggregating data weekly. To allow a global comparison on a common time range (seven years), the concurrently quoted CALs have been averaged in one single time series. We notice positive correlation between the year-ahead prices and all considered time series with the exception of the two stock indexes (S&P350 and Bel20).

The Granger causality makes the assumption of stationary data. We thus replace the time series by the one step percent change and obtain the results displayed in Figure (5.13). Reading the figure is done as follows: if we take the value 0.1 in (row 2, column 1), it refers to the lowest p-value of the Granger test for Mean CAL (X) causing SXEP (Y). The lag $k$ which leaded to this p-value is given by the corresponding cell in the rightmost heatmap. A p-value lower than 0.05 reject the null hypothesis *X do not cause Granger Cause Y*.

According to the statistical test, the time series tend to interchangeably causing each other. Since the dataset features seven years of common data, this is not much surprising. LPG and fuel oils are not expected to improve the forecast of year-ahead prices. However, several indications show the results don't seem very much reliable. The best lag value $k$ is often encountered with an unexpected high lag. In this way, it is highly unexpected that a lag value of 11 would be the most advantageous. Furthermore, the Granger test is not coherent with the previous findings. From Figure (5.12), we noticed that the stock indexes weren't very correlated with the year-ahead prices which was then loosely confirmed by a manual graphical inspection. Meanwhile, the Granger test insures that the two indexes were relevant with high statistical significance. We conclude that the Granger Causality likely doesn't provide reliable information about causality between the related time series.

Overall, the empirical tests in this study highlight a positive correlation between the year-ahead electricity prices, other commodity prices and energy indexes but fails to find additional compelling and consistent evidence of predictive links. Still, the restropective analysis study of the M4 competition[50] recommends to "*select variables in forecasting with loose significance*". In addition, the deep learning models we reviewed in Chapter 4 have the benefit to learn the dynamics of many time series automatically[4], and, therefore, prevent the need of more labor intensive correlation studies in the dataset.

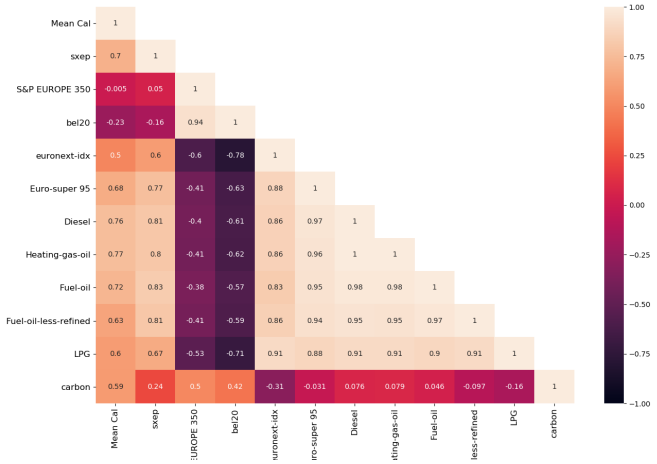
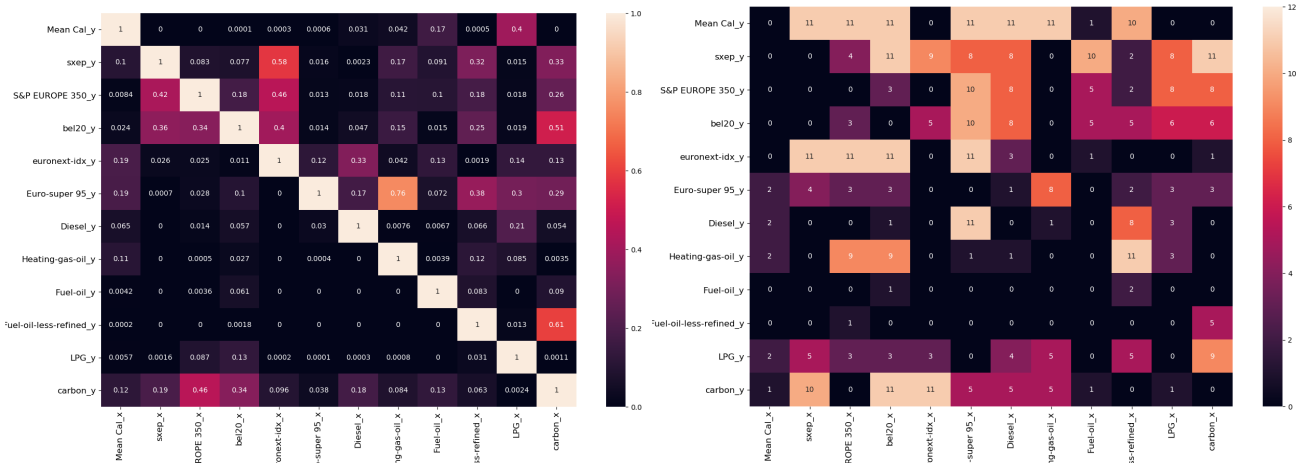



Figure 5.12 – Spearman correlation between all scaled variables ([0, 1])




(a) Lowest p-values for lag $k \in \{1, \ldots, 12\}$.

(b) Lag k corresponding to the lowest p-value.

Figure 5.13 – Summary of the Granger Causality test. The leftmost figure gives the lowest p-value obtained for a lag $k$. If $p < 0.005$, the statistical test states X Granger causes Y.

# 6 | Forecasting

Quantile forecasts are central to risk management decisions because of the widespread use of Value-at-Risk (VaR). In this chapter, we first explain the methodology used to forecast the YAM prices with such methods as well as the performance assessment methodology. We then expose the empirical results obtained and provide a discussion about them. Suggestions for improvements are expressed before finally giving a conclusion on the results usability for reducing cost of electricity acquisition.

## 1 Methodology

### 1.1 Data preparation and pre-processing

#### 1.1.1 Log transformation

The *log* transformation is a frequently used pre-processing technique for forecasting tasks. Helmut et al.[80] gives an in-depth overview of the advantages and disadvantages of such transformation on economic variables with simple linear forecast techniques. Their results show that using the log transformation can produce dramatic gains in forecast accuracy if it makes the variance more homogeneous through the sample. Otherwise, the pre-processing might reduce accuracy but to a lower extent. On stock indexes, the transformation turns out to be on average beneficial except for the longest forecasting periods. Their interpretation is that long periods sometimes cover general downturns in stock markets, which penalized the forecast accuracy of log transformed data in these particular cases.

Figure (6.1) reports the evolution of the variance across the CAL 2016 price series with a window size of four weeks. Since ordinary values, and therefore the variance, are obviously affected by the log transformation, the resulting series have been scaled to a common range. It is observed that the transformation seems to slightly stabilize the variance. Yet, a look at Appendix 5.3.2 where the same plots have been derived for CAL 2013 and CAL 2019 shows that this result is not consistent. Indeed, the variance for CAL 2013 didn't change much at all. Conversely, the one for CAL 2019 was negatively impacted by the additional pre-processing. Mixed results are thus obtained *restrospectively*. In this sense, it should be clearly understood, that, in practice though, the forecaster wouldn't be able to assess the benefits of the transformation *a-priori*. Still, since many practitioners[81, 82] take the log-transformation for granted, we include this processing step in our data preparation phase.

#### 1.1.2 Scaling

It is now widely accepted that inputs for neural-network based techniques should be shifted and scaled so that their average is somewhat close to zero[83]. Instead of a hazardous data standardization which would change the shape of the input distribution, we simply scale the training data in $[0, 1]$ so that the shape of the input distribution is not affected. Scaling an arbitrary input vector $Z \in \mathbb{R}^{1\times n}, n > 0$ results in the transformed vector $Z' \in \mathbb{R}^{1\times n}$:

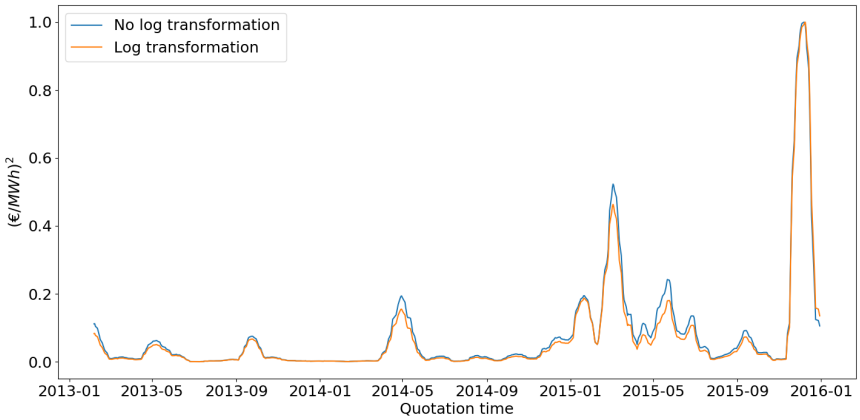



Figure 6.1 – Comparison of the scaled rolling variance evolution on raw and log-transformed prices of CAL 2016. The window size is fixed to $|w| = 4 \times 7 = 28$ days.

$$Z' = \frac{(Z - min(Z))}{(max(Z) - min(Z))} \tag{6.1}$$

with $z'_i \in [0,1], i = 0,\ldots,n$. Let's note this mapping can be generalized to scale the input in any specified interval $[L,U], L \in \mathbb{R}, U \in \mathbb{R}$:

$$Z'' = Z'.(U - L) + L \tag{6.2}$$

### 1.1.3 Weekly data aggregation

In order to compare the forecast accuracy obtained with univariate and multi-variate predictions but also to leverage these exogenous variables, we need to aggregate the prices weekly so that the sample frequency is the same in both settings. This is not without consequences. Indeed, it results in shorter training sets which are likely to deteriorate the forecast accuracy on early open days of a given CAL. Furthermore, we loose weekday granularity in the forecast. Considering we are however interested in medium to long term trends, the later is not really a problem. Aggregating the data also enables to counterbalance the loss of training size with a reduction of the forecast horizon. Furthermore, when aggregating the data weekly, missing values corresponding to closed market days are simply ignored, which removes possible discussion on regression techniques to fill them. No smoothing technique is applied before the resampling.

## 1.2 Exogenous variables and concurrently quoted CAL

In view of the exogenous variable analysis (Subsection 3.1) and the more restricted time range available in the dataset, stock index data corresponding to Bel20 and S&P350 are removed. The case of concurrently quoted CALs is interesting. Only three CALs are quoted at the same time and one CAL has a quotation period that lasts over three years. As a result, by focusing on the predictions of one CAL, we can't use the same triplet all along. Let's for instance consider the case of CAL 2016 as an illustration. CAL 2014 up to CAL 2018 have actually a common quotation period with CAL 2016 at some time or another. Including all of them in the training data would require to introduce zero-padding which might be harmful during model training. To tackle this problem, we only include other concurrently quoted CAL data, up to the forecast start date, if their quotation time can be exactly mapped one to one with the CAL being forecast. This mechanism is visually depicted in Figure (8.15) for three forecast start scenarios in Appendix 6.0.1.

## 1.3 Models configuration

In light of the literature review and data analysis carried out earlier, we consider a diversified pool of models sharing the same API in order to quickly evaluate a set of diversified forecasting techniques. Thereby, we leverage the recent gluon-ts [53] library which specializes in prediction interval forecasting and evaluate the results on all the (working) models of the API. The models considered are the one presented in Section 2 with the addition of Gaussian Processes[84] and DeepFactor[85] as additional benchmarks.

**Hyperparameters tuning** All the algorithms are kept with their *default* hyperparameters without further grid search, to keep the computation cost under control. Only DeepState and DeepAR have seen their number of epochs shrunk to 5 and 20 respectively due to their very slow training phase.

**Forecast horizon** In Subsection 1.2, we assumed two different restrictions on the minimum block power size (`block_size`) which translated in 10 or 50 purchase decisions to dispatch across the quotation period for a required base load of $10MW$. Assuming 3.365 = 1095 open market days, splitting uniformly the quotation period in 50 trading sub-periods as proposed for the baseline hedging algorithm (Subsection 1.1) would create a time gap between two purchase decisions of $\frac{3.365}{50} = 21.9$ days. In this way, when a new time slot begins and time comes to decide whether it is better to buy electricity $now$ or in the next time slot, a forecast of approximately three weeks is required. As pointed out in Makridakis et al.[47], the prediction intervals are likely to underestimate reality and this underestimation will increase as the forecasting horizon lenghtens. This motivates us to constrain at first the forecast horizon to four steps and not more. Making a reliable forecast of $\frac{3.365}{7.10} \approx 15$ weeks, for a 10MW `block_size` constraints, is more than likely overly optimistic for a first approach.

**Prediction interval** Both forecast competitions introduced in Chapter 4 took different approaches. On the one hand, the GefCom2014 asked competitors to produce the 99 quantiles $[q_{01}, q_{02}, \ldots, q_{99}]$ as a discrete approximation of the forecast distribution. On the other hand, the $M4$ competition took inspiration from the value-at-risk framework and only required competitor to produce the 95% (confidence level of $\alpha = 5\%$) prediction interval built from the quantiles $q_{97.5}$ and $q_{2.5\%}$. For our experiments, all the models are configured to produce the $q_{01}$ and $q_{99}$ quantiles so as to generate a 98% prediction interval. All the models relying on Monte Carlo sampling are set to produce 100 samples (as default) to derive these quantiles.

## 1.4 Model performance evaluation

### 1.4.1 Considered quotation periods

To ensure that the results are not driven purely by specific periods, we report the results for a quotation period spanning from 2010 to 2018 both inclusive. Considering that the rolling evaluation pipeline, presented in Subsubsection 1.4.2, requires a lot of computational resources, a trade-off to achieve statistical significance, we limit the evaluation to CAL 2013, CAL 2016 and CAL 2019. In this way, we ensure that we indeed cover independent time series regimes since the quotation of these CALs never overlaps. Indeed, Concurrently quoted CALs tend to show redundant patterns (Subsection 3.2) anyway. Furthermore this selection provides a good diversity of price distributions (Figure (5.2)). CAL 2013 is among the least volatile CAL and the prices distribution is nearly symmetrical with regards to the median. CAL 2016 has already been discussed extensively and CAL 2019 would allow us to analyze the forecast behavior on an exponentially increasing trend, whereby the maximum prices (72.12€/MWh) account for a soaring 266% of the minimum value (27.1€/MWh) observed in the same CAL: the highest difference ever observed for the complete dataset.

### 1.4.2 Backtesting pipeline

*Cross-validation* is a popular procedure for tuning hyperparameters, selecting models and producing robust measurements of models performance that are statistically significant. Due to the predominant temporal dependencies in the dataset, conventional techniques like *k-fold* or *leave-one-out* would introduce data leakages.

**Rolling origin evaluation** Instead, we consider an evaluation on a *rolling origin* principle[7, 86]. Let the forecast origin be the time point of the last known value, from which the forecast is performed. The underlying principle of the procedure consists in performing a set of fixed horizon forecasts by sequentially moving values from the test set to the train set and changing the forecast origin accordingly[87]. The forecast origin thus "rolls forward in time" as shown in Figure (6.2) for a simple univariate two-step forecast. Considering the exogenous variables expose positive correlation, the procedure needs to be applied on all the time series simultaneously. A complete retraining of the model is required to adjust the model to the newly observed values. To avoid any look-ahead bias, the transformations exposed in Subsection 1.1 are applied each time on the training data window. Coupled with the chosen forecast horizon, this pipeline turns out to require 35 simulations to cover the quotation period of each considered CAL.

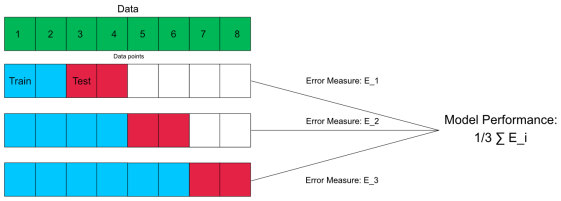



Figure 6.2 – *Rolling origin* backtesting scheme for a *univariate* time series and a forecast horizon of two.

**Extension for model selection** The inclusion of an inner loop for *a-priori* model selection is a natural extension of the previously describe backtesting strategy. This extension is sometimes referred as *rolling-origin-recalibration*[87] or *forward-chaining nested cross-validation*[7]. As illustrated in Figure (6.3), models are first trained on a window and their performance is assessed on the following validation window. The most accurate model is selected and then trained again by assimilating the validation window in the training set. The reported error measures are the one of the test window. Since some models might be better adapted to particular local regimes of the time series, performing a model selection is an experiment worth considering. Likewise, model parameters could be optimized with this scheme but won't be considered to keep the computation cost under control.

It should finally be noted that the default parameters of the Gluon-ts library may limit the size of the training set automatically to a fixed window whose length simply equals the forecast horizon for some estimators.

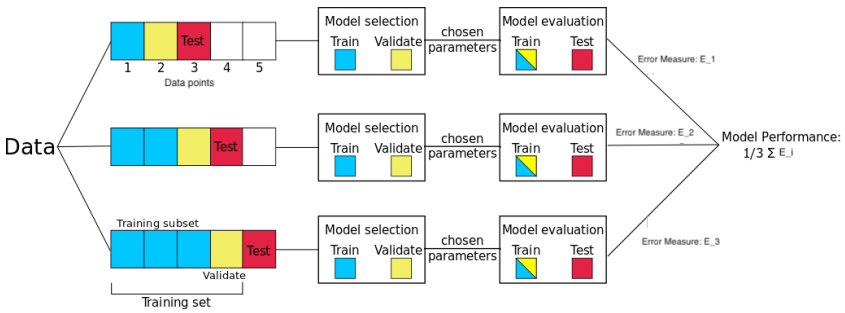



Figure 6.3 – *Rolling origin with recalibration* backtesting scheme for a *univariate* time series and a one step forecast[7].

### 1.4.3 Error measures

For a given forecast origin, a set of error measures averaged over all forecast points are computed for each model. We then only report, for each model, the average of these intermediate measures over all the simulations on the *test set* by being aware that this might penalize some models that would perform well most of the time but underperform in some rare specific situations though. The measurement is naturally performed between the inverse scaled prediction and the actual values. For a given fixed number of forecast step H, let's denote by **y** a generic time series which take values $y_t \in \mathbb{R}, t = 1, \ldots, T + H$ with forecast $\hat{y}_t$ produced over $T + 1, \ldots, T + H$.

To assess the quality of an estimator, we not only report four error measures, but choose these measures to provide complementary interpretation of the results. In this way, we first report the widely used [46] *Mean Squared Error* which naturally measures the average squared difference between the estimated values and the actual values:

$$\text{MSE} = \frac{1}{n} \sum_{t=T+1}^{T+H} \left(y_t - \hat{y}_t\right)^2 \tag{6.3}$$

It is a scale dependent, yet simple non-negative measure, which, in the case of an unbiaised estimator, fully reflects the variance and irreducible uncertainty of the estimator. It can't however be used to compare the performance across different CAL. At the cost of a more straightforward interpretation like the mean absolute error, the introduced square exponent has the advantage to greatly penalize large errors.

Another error measure that became quite popular with the advent of forecasting competitions [47] is the Mean Absolute Scaled Error first introduced by [88]. This error metric belongs to the family of scale free error measures since it gives each error as a ratio compared to a baseline average error. For instance, the conventional MASE is defined as the mean absolute error of the forecast values, divided by the mean absolute error of the in-sample one-step naive random walk forecast[88, 46]:

$$\text{MASE} = \frac{\frac{1}{H} \sum_{t=T+1}^{T+H} |e_t|}{\frac{1}{T-1} \sum_{t=2}^{T} \left|y_t - y_{t-1}\right|} \tag{6.4}$$

$$\tag{6.5}$$

where $e_t = y_t - \hat{y}_t$ denotes the mean absolute error of the forecast for period $t$. [88] recommends its use as a versatile benchmarking tool, superior to the overly used and unstable (with actual values close to zero) symmetric mean absolute percentage error (sMAPE) [89].

The next error measure aims at specifically evaluating prediction intervals. As prediction intervals slowly made their introduction in literature, it doesn't seem to be a true consensus on a PI scoring metric but the Mean Scaled Interval Score (MSIS), first introduced by [90] has been at least used in the M4 competition [49]. This quality measure aims at balancing coverage and width of prediction intervals since we strive for a tradeoff between high coverage but with short, non-trivial, intervals so that they are at least informative. MSIS is defined as follows:

$$\text{MSIS} = \frac{\frac{1}{H} \sum_{t=T+1}^{T+H} \left(\hat{U}_t - \hat{L}_t\right) + \frac{2}{a}\left(\hat{L}_t - y_t\right) \mathbb{1}\{y_t < \hat{L}_t\} + \frac{2}{a}\left(y_t - \hat{U}_t\right) \mathbb{1}\{y_t > \hat{U}_t\}}{\frac{1}{T-m} \sum_{t=m+1}^{T} \left|y_t - y_{t-m}\right|} \tag{6.6}$$

where $\hat{L} \in [0,1], \hat{U} \in [0,1], \hat{L} < \hat{U}$ are the lower and upper bounds of the prediction, y the future observations and $\mathbb{1}$ the indicator function. The significance level $\alpha$ is then used to control the target interval width ($\alpha = 0.005$ for 95% PI for example). The denominator scales the interval score by the in-sample seasonal error. In the ideal case, the first term would be close to 0 whereas the two next penalizing terms would be exactly 0 if all the actual values fall correctly within the prediction interval bounds. The M4 competition guideline report[49] gives an additional empirical example to interpret the penalizing terms. The complexity of such measure makes it already hard to interpret in a stand alone manner though. For this reason,

we won't include other more complex scoring rules which have been developed in the literature to evaluate probabilistic forecasting [45].

Finally, we include an additional simple qualitative indicator. The Prediction Interval Coverage Probability[46, 45] (PICP) summarizes the proportion of actual values that fall inside the prediction interval. By definition, a prediction interval $[\hat{L}_t, \hat{U}_t]$ with level $(1-\alpha)$ should match the nominal rate $P(y_t \in [\hat{L}_t, \hat{U}_t] = (1-\alpha))$. If the condition is met, the later is then said to be *calibrated*. PI forecasting obviously strives to generate calibrated but also *sharp* (narrow) informative quantiles. While providing a straightforward way to interpret the calibration, the PICP alone doesn't provide any information about *sharpness*, in opposition to MSIS. For these reasons and despite its simplicity, PICP well complements MSIS. Let's note the same principle can naturally be used to compute any quantile coverage by measuring the number of observations that fall below a each of the desired quantiles.

### 1.4.4 Naive predictor benchmarking

To benchmark the different models of Gluon-ts, we compare them with two different naive predictors. We consider the *naive random walk* forecast which simply predicts the last observed value all along the prediction range. By using the same notation as in Subsubsection 1.4.3, the model outputs at forecast time $T$: $\hat{y}_t = y_T,\ \forall\, t \in \{T+1, \ldots, T+H\}$. We refer to this model as *Trivial RW* later on. The second naive predictor generates K forecast samples whose value at each time $t$ in the prediction range is given by $\tilde{y}_t^k = \mu + s.\sigma$ with $s \sim \mathcal{N}(0,1)$ and where $\mu$ and $\sigma$ are the mean and standard deviation of the last N observed values. We consider $N = 2$ and stick to the Gluon-ts terminology by calling this model *Trivial Mean 2*.

## 2 Results for univariate time series prediction

In order to evaluate empirically and *a-posteriori* if adding exogenous variable is beneficial, we first report the results obtained by training all the estimators without adding the exogenous variables.

The total computation time to run the experiment on three different CALs is estimated to twelve hours on an intel i9 9880H with 8 threads dedicated to the computation. Most of the computation has however been carried out on the provided CECI cluster infrastructure (NIC4[91]) while DeepState, MQRNN, MQCNN and Gaussian Processes had to be run on Google Colab[92]. The simulation on Colab took eight hours to complete for the three CALs with GPU support activated for MQCNN and MQRNN. Thanks to jobs parralelization, the NIC4 cluster was able to handle one forecast simulation ($3 \times 35$ of them) in 20 minutes approximately (queuing time excluded). Table 6.1 summarizes the results obtained for a *rolling origin* evaluation and a forecast horizon of four weeks.

**Trivial predictors** The bottom of Table 6.1 displays the naive predictors introduced in the previous section. By paying particular attention to *Trivial RW*, it can be noticed at a glance that all the models without exception perform worse. In fact, while being the simplest model, it turns out to outperform all the more complex estimators by a good margin. In this way, the empirical results of Table 6.1 show that complex models can't provide additional accuracy on the random-like time series of the year-ahead market. In this sense, it indeed confirms the findings of Chapter 5. Looking at *Trivial Mean 2*, we notice it performs second overall for most accuracy measures. The historical sampling of NPTS, even though performed on a relatively short past window of 4 weeks, didn't provide better results.

**Non-trivial estimators** Among the most complex models, *DeepState* produces the safest prediction intervals. While the reported results for the other evaluation metrics are disappointing, it is at least valuable that the model doesn't produce overly optimistic prediction intervals as assessed by the PICP. In this sense, the model successfully reports a greater uncertainty on the future price distribution and is aware of the forecast difficulty. MSIS naturally takes this property into account and therefore ranks *DeepState* higher. For the same reason, the baseline regression-inspired *Prophet* scores among the best in terms of MSIS. In terms

| | Cal 2013 | | | | Cal 2016 | | | | Cal 2019 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | MSE | MASE | MSIS | PICP | MSE | MASE | MSIS | PICP | MSE | MASE | MSIS | PICP |
| **DeepAR** | 4.87 | 2.04 | 63.50 | 0.14 | 4.18 | 3.72 | 114.80 | 0.35 | 14.63 | 3.82 | 108.35 | 0.24 |
| **DeepState** | 13.97 | 3.74 | 40.27 | 0.72 | 6.93 | 5.87 | 60.23 | **0.83** | 15.72 | 4.85 | **47.58** | 0.70 |
| **DeepFactor** | 213.57 | 4.39 | 175.02 | 0.00 | 6.80 | 5.62 | 223.75 | 0.00 | 17.05 | 4.64 | 184.77 | 0.00 |
| **MQCNN** | NaN | 2.09 | 72.70 | 0.06 | NaN | 4.08 | 144.47 | 0.10 | NaN | 3.07 | 104.49 | 0.08 |
| **MQRNN** | NaN | 1.99 | 80.52 | 0.06 | NaN | 3.51 | 125.58 | 0.10 | NaN | **2.62** | 85.50 | 0.08 |
| **WaveNet** | 60.62 | 7.26 | 86.15 | 0.43 | 16.50 | 9.15 | 128.89 | 0.46 | 136.60 | 11.08 | 135.39 | 0.48 |
| **Prophet** | 10.21 | 3.18 | 29.20 | 0.84 | 6.13 | 5.85 | 61.34 | 0.81 | 10.37 | 3.89 | 49.11 | 0.78 |
| **Gaussian Processes** | 4.68 | 1.88 | 32.22 | 0.41 | 3.84 | 3.68 | 65.07 | 0.42 | 16.74 | 4.73 | 119.00 | 0.28 |
| **NPTS** | 15.87 | 4.42 | 106.30 | 0.32 | 8.65 | 7.93 | 194.70 | 0.31 | 108.09 | 11.80 | 379.03 | 0.09 |
| **FFNN** | **2.35** | **1.33** | **21.49** | **0.85** | **2.09** | **2.57** | **48.52** | 0.80 | $6 \times 10^9$ | 2.71 | 60.83 | **0.94** |
| **Trivial RW** | **1.81** | **1.24** | / | / | **1.67** | **2.44** | / | / | **5.49** | **2.48** | / | / |
| **Trivial Mean 2** | 2.25 | 1.40 | 32.31 | 0.31 | 2.09 | 2.86 | 64.39 | 0.40 | 6.18 | 2.67 | 70.03 | 0.25 |

Table 6.1 – Prediction accuracy assessment of univariate forecasting on three different CAL. Backtesting carried out with of a rolling origin evaluation and a time horizon of four weeks. The Bottom of the table display the naïve models performance.

of MSE, it can be noticed that Gaussian processes, which are more suitable to stationary data, struggle to handle the increasing trend of CAL 2019. Wavenet also produces worse results on this CAL. Interestingly, the simple feedforward neural network is the "best performing" model among the non-trival ones. Eventhough its wide quantile predictions are likely not very much informative, the model produces safe prediction intervals and has a comparable accuracy for the median prediction as Trivial Mean 2. The FFNN strives the best balance between coverage and median prediction accuracy among non-trivial models. It is likely that it produced a few completely off predictions on CAL 2019 though, as can be interpreted from the very large MSE. The *NaN* entries for the MSE of MQCNN and MQRNN are both a result of an unbounded MSE produced for the first simulation, whereby the training set length exactly equals the prediction length.

# 3 Results for multi-variate time series prediction

## 3.1 Model performance evaluation and discussion

Table 6.2 summarizes the results obtained by including 10 additional exogenous variables as discussed previously. The table features two additional entries: an additional one for DeepAR and another one for the Feed Forward Neural Network. They are actually the same model but trained on the raw scaled data rather than the log scaled data. Considering they were the two best performing models, the experiment without the log transformation has been only repeated for those two. Compared to the univariate experiment, computation time didn't change much on the NIC4 cluster and increased by less than 10% on Google Colab which might as well be inherent to the share of hardware resources among users.

**PIs Calibration improvements** A first observation that can be made is that including the exogenous variable turned out to substantially improve the calibration of the prediction interval produced by all the models with the little exception of DeepState. In this way, almost all the models gained better knowledge on the range of future price outcomes.

**Forecast accuracy improvements** With regards to the other error measures, we observe a noticeable gain in accuracy too. Yet again, DeepState did worse than previously though. A possible explanation is that, since DeepState is by far the most time consuming model to train, the training error might not have had the time to stabilize due to the restricted number of epochs. Moreover, we didn't include any additional features

| | | Cal 2013 | | | | Cal 2016 | | | | Cal 2019 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **Log** | MSE | MASE | MSIS | PICP | MSE | MASE | MSIS | PICP | MSE | MASE | MSIS | PICP |
| **DeepAR** | ✓ | 3.70 | 1.76 | 28.94 | 0.64 | 2.32 | 2.88 | 43.51 | 0.64 | 6.65 | 2.53 | 36.22 | 0.68 |
| **DeepAR** | × | 4.74 | 2.02 | 34.58 | 0.54 | 3.18 | 3.35 | 51.85 | 0.60 | **5.92** | 2.55 | **36.04** | 0.64 |
| **DeepState** | ✓ | 16.61 | 4.43 | 59.02 | 0.61 | 6.11 | 5.85 | 61.88 | 0.79 | 27.07 | 5.76 | 64.06 | 0.80 |
| **DeepFactor** | ✓ | 25.42 | 5.98 | 32.33 | 0.70 | 6.87 | 6.32 | 45.94 | **0.89** | inf | inf | 52.49 | 0.66 |
| **MQCNN** | ✓ | NaN | 2.49 | 53.46 | 0.40 | NaN | 3.99 | 104.23 | 0.30 | NaN | **2.37** | 57.09 | 0.40 |
| **MQRNN** | ✓ | NaN | 2.62 | 59.82 | 0.28 | NaN | 3.80 | 101.32 | 0.33 | NaN | 2.46 | 57.35 | 0.35 |
| **WaveNet** | ✓ | 43.95 | 6.20 | 49.47 | 0.71 | 4.56 | 4.27 | 45.65 | 0.75 | 15.42 | 4.23 | 50.22 | 0.81 |
| **FFNN** | ✓ | 1.85 | **1.26** | **20.51** | 0.84 | **1.86** | **2.49** | **46.29** | 0.85 | 11.25 | 2.47 | 49.13 | 0.95 |
| **FFNN** | × | **1.72** | **1.26** | 20.93 | **0.85** | **1.71** | **2.44** | **45.79** | 0.84 | 5.98 | 2.52 | 40.96 | **0.97** |
| **Trivial RW** | / | **1.81** | 1.24 | / | / | **1.67** | **2.44** | / | / | **5.49** | **2.48** | / | / |
| **Trivial Mean 2** | ✓ | 2.25 | 1.40 | 32.31 | 0.31 | 2.09 | 2.86 | 64.39 | 0.40 | 6.18 | 2.67 | 70.03 | 0.25 |

Table 6.2 – Prediction accuracy assessment of multivariate forecasting on three different CAL. Backtesting carried out with of a rolling origin evaluation and a time horizon of four weeks. The log column makes reference to the use of logged transformed prices for training.

for each related time series while DeepState relies primarily on them to learn a global model. Interestingly, MQCNN and MQRNN score a lower MASE on CAL 2013 than previously but produce better calibrated PIs on the other hand. A look at the MSIS actually shows that the exogenous variables should be considered beneficial. Another observation is that previous abnormally high error scores encountered in FNN, Wavenet and Deep Factor got significantly reduced.

**Comparison with univariate baselines** Gaussian processes and Prophet are not displayed in Table 6.2 since they don't support multivariate predictions. However, looking back at Table 6.1, it can be noticed that only DeepAR and the FFNN do a better score consistently. MQRNN and MQCNN have a lower MASE than Prophet, all the CALs considered but have a higher MSIS due to worse PIs calibration. MQCNN, MQRNN and Wavenet outperform Gaussian Processes on CAL 2019.

**Best performing multivariate models** Among the non-trivial multivariate models, the simple feedforward neural network turned out to produce the best price predictions on CAL 2013 and CAL 2016. On CAL 2019 though, DeepAR takes a little hedge over the simpler neural network with a twice lower MSE and a better MSIS. However, it also produces overly confident prediction interval with a PICP of only 64% instead of the target 98%. Meanwhile, the FFNN has nearly perfectly calibrated (but likely not sharp) PIs with a PICP of 95% on average.

**Impact of the log transformation** The same experiment repeated for the two best performing models, DeepAR and FFNN, with the data log transformation removed, shows the transformation doesn't impact the two models in exactly the same way. The FFNN saw prediction accuracy improvements by removing the transformation on each CAL. The most notable improvements are on the exponentially increasing CAL 2019 where the MSE gets almost halved. On the other hand, DeepAR took advantage of the log transformation on CAL 2013 and CAL 2016 but, like the FFNN, better results are obtained without the log transformation on CAL 2019. In this sense, the results of this additional experiment are in line with the discussion in Subsubsection 1.1.1. Since the preprocessing doesn't provide consistent benefits retrospectively, the log preprocessing should probably be discarded though.

**Comparison with univariate trivial predictors** As it turns out, no model is able to undeniably produce better results than the naive forecast despite the data augmentation. However, the FFNN shows very similar performances compared to the naive benchmark on all the CALs with even a better MSE for CAL 2013 and a

better MASE on CAL 2019 for the log version. The neural network is also able to produce very well calibrated prediction intervals while the naive predictor doesn't produce any by definition. In fact, Trivial RW may not be the optimal model among the considered pool on CAL 2019. DeepAR produces close results to Trivial RW too while MQCNN and MQRNN have even a lower MASE than the trivial predictor. MQCNN and MQRNN are however underperforming in terms of MSIS by being penalized for their bad PI calibration. Still, all the CALs considered, trivial RW has still a very slight edge on the median prediction accuracy over the FFNN despite not providing any information about the future price ranges.

**Impact of model selection** Up to now, results have been discussed for a *rolling origin* evaluation without any *a priori* model selection on a validation window. Table 6.3 summarizes the results with the *rolling origin with recalibration* evaluation scheme presented in Subsubsection 1.4.2. Models were selected on the basis of the MASE score. The resulting sets of forecasts didn't yield lower error scores. Trivial RW, FFNN and DeepAr were overall the most often selected models but most models got included at east once in the end. A good accuracy on the validation set is thus not a good enough indicator of reliability since some models likely turned out to be selected by luck and drove the accuracy down.

| | Cal 2013 | | | | Cal 2016 | | | | Cal 2019 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MSE | MASE | MSIS | PICP | MSE | MASE | MSIS | PICP | MSE | MASE | MSIS | PICP |
| **Model selection** | 2.453 | 1.367 | 24.347 | 0.614 | 4.178 | 3.843 | 63.557 | 0.636 | 13.255 | 2.904 | 49.700 | 0.632 |

Table 6.3 – Prediction accuracy summary on the test set on a rolling origin with recalibration pipeline which select the best model. Both the validation window and the testing window length are set to four weeks.

## 3.2 Forecast visualization and residuals

The FFNN simulations were characterized by a 5% lower MSE compared to Trivial RW on CAL 2013. The following paragraphs give an additional analysis on the results obtained with the feed forward neural network on this CAL.

From Figure (6.4), we notice the model didn't tried to systematically reproduce the flat predictions of the benchmark but started to do so around the beginning of 2012. When prices tends to fluctuate more, the prediction interval correctly widens. The model produced overly optimistic prediction intervals during 2012.

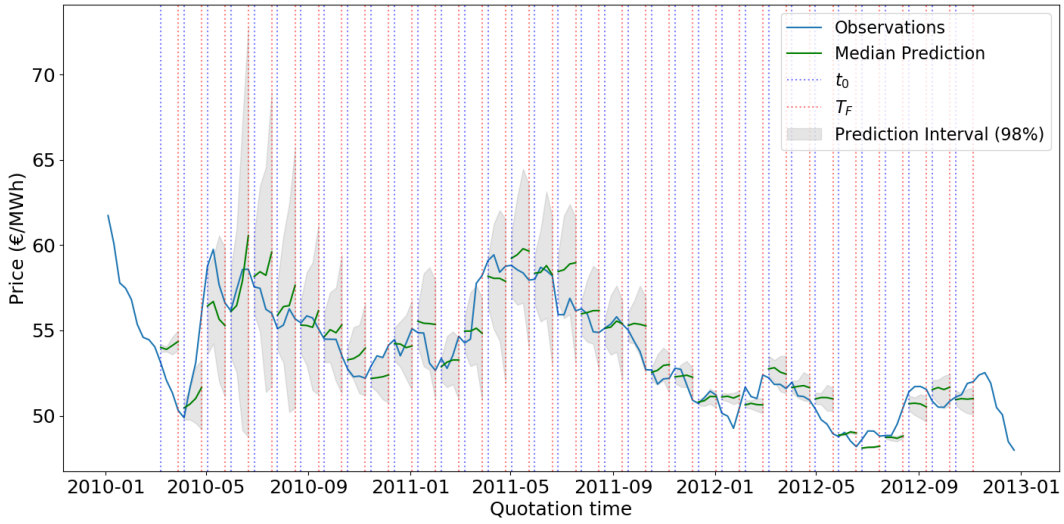
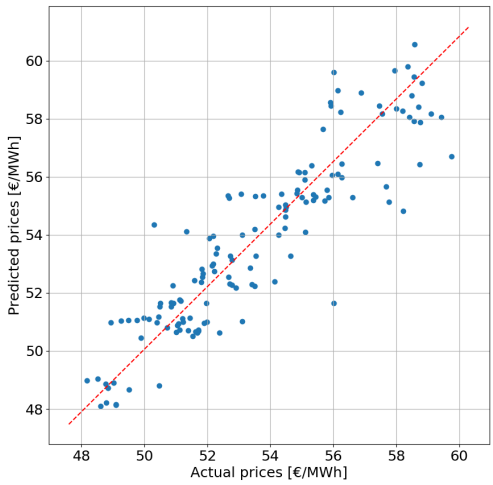


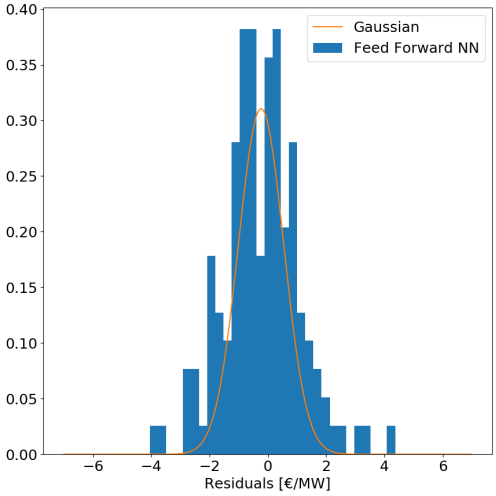

Figure 6.4 – Feed forward neural network predictions for CAL 2013

The FFNN yields an average difference between predicted prices and actual prices (denoted *residuals*[46] from now on) of $-0.23$€/MW. In this way, the estimator is rather *unbiaised*, at least on CAL 2013 which, pointed out earlier, has already a symmetric price distribution. From Figure (6.5a), which displays all the predicted points, we notice that the predictions tend to be centered around the red line. The closer the points to the later the better. Higher prices tend to deviate more though.

The histogram in Figure (6.5b) shows residuals also tend to follow a normal distribution. This observation is likewise confirmed by the corresponding QQ plot, Figure (8.16), in Appendix 6.0.2. A look at the horizontal axis of the histogram tells us the amplitude of residuals are still non-negligible though. For instance, the maximum difference between the predicted and actual prices amounts to 4.364€/MW. In terms of monetary value, a missjudgment like so already represents approximately a potential loss of $4.364 \times 60 \times 24 = 38,230$£ assuming just 1MW has been purchased.




(a) Scatter plot of the predicted prices against ground truth.




(b) Residual diagnosis.

Figure 6.5 – FFNN visual forecast evaluation on CAL 2013.

Finally, it is worth quickly investigating whether the model can be used to correctly predict simple price directions. To do so, we only consider a very simple assessment procedure based on the median point-wise prediction. In this sense, more complex evaluations leveraging the predicted future price distribution should be encouraged instead.

A confusion matrix can be for example computed by categorizing all the point-wise median predictions into two categories: either *up* or *down*. Any prediction whose value is greater (less) than the last observed price, from which the forecast is generated, is labelized as *up* (*down*). The same procedure is applied in a similar way to actual values. From the resulting Table 6.4, we notice that the FFNN struggles to predict the correct price movement accurately, at least by considering the median as a proxy of the price direction. When actual prices went down, the model was able to correctly predict the direction most of the time. However, when prices went up, the model completely failed to capture the correct direction since the model predicted 51 times the wrong direction and only 38 times the correct one. From a practical point of view, a company energy manager who would have relied on the forecast to evaluate whether it was opportune to buy in this situation, would have likely received a positive signal. If the manager indeed followed the indication and purchased a block order, he would have therefore paid the higher price. As a result, the median point-wise prediction should not be considered as a reliable indicator of future price direction, at least from a restrospective point of view on CAL 2013 with FFNN.

| **Actual** \ **Predicted** | Down | Up |
|---|---|---|
| Down | 49 | 6 |
| Up | 51 | 38 |

Table 6.4 – Confusion matrix derived from the prediction of FFNN on CAL 2013.

# 7 | Conclusion and perspectives

We began this Master thesis by reviewing the current European electrical system from three different perspectives. First, we briefly exposed the notion of grid infrastructure, introduced the main actors and highlighted some trends in the Belgian system. We then gave a comprehensive overview of the markets available to trade electrical energy and highlighted some of their properties. Lastly, a trading focus was adopted. We deduced from a delivery-wise price comparison that trading in the year-ahead market largely reduces the exposure to the volatile prices encountered in the spot market and, in that sense, turns out to be a good strategy to hedge against risk. All those introductory notions led us to discuss several power price drivers where we learned how complex price dynamics can be.

The analysis aimed at improving the decision making process of a rather risk-averse industrial company looking to reduce its cost of electricity acquisition by interacting in the year-ahead market. More specifically, it was targeted to help such company securing a fixed base load supply of around 10MW using a generic *click-by-click* contract type. Specific constraints included a limited number of orders to dispatch throughout each quotation period and the expectation that the research outcome should provide an indicator to complement the expertise and autonomy of the resource manager. The absence of relevant support to the problem in the literature motivated us to embrace the probabilistic forecasting approach to provide a flexible support to human decision making while providing nuanced information over future price outcomes. To this end, we planned to leverage recent advances in deep learning based models and reviewed some of the most prominent models featured in the literature in the last four years.

To expand our knowledge of the price dynamics ruling the year-ahead market, an extensive data analysis study was conducted on the available dataset. It showed that the price distributions were not consistent from a delivery year to another and often highly skewed. The most noticeable finding of the analysis was that the year-ahead electricity prices exposed very similar properties to random walk processes whose best theoretical prediction is simply the last observed value. In lights of the reflection on power price drivers and the literature review, we decided to analyze several related times series datasets fetched proactively from the web to mitigate this problem. Similarity measures and the Granger Causality were the two main techniques to analyze correlation and causality among eleven time series as it is common in finance. Seven commodities, two energy indexes and two stock indexes were considered. It was shown all time series except the two stock indexes are positively correlated with the year-ahead prices on a seven years long time frame. Mixed results were obtained with the Granger Causality though. We suggest to not rely on this tool for future work due to the difficulty to interpret the results and evaluate their reliability.

By means of a rolling origin backtesting methodology, we assessed the results of both univariate and multivariate probabilistic predictions on the year-ahead market. Ten models were evaluated for the univariate setup and seven for the multivariate one on cloud computing servers. The results are clear. The time series model yielding the lowest prediction errors collapses to a simple and non-informative constant point-wise prediction generated from the last observed price: the naive random walk forecast. The later doesn't provide much information to measure opportunity. Despite leveraging other related time series, the best performing (non-trivial) model is a simple feed forward neural network that produces very similar results to the random walk benchmark when exogenous variables are used in the training set. However, its prediction quantiles can be used to estimate future price ranges with a decent accuracy. In this sense, the neural network provides additional benefits over the benchmark although not outperforming it in a consistent way in terms of median prediction accuracy.

While common belief would have gone in favor of complex deep learning based time series models, our results are still coherent with the finding of the M4 competition[50] whereby simpler models often tend to perform very well comparatively. Naturally, our experiments don't pretend to have covered all the possible ways to predict the future price outcomes. In this way, this Master thesis only considered a continuous probabilistic forecasting approach. Threshold forecasting, price bins classification, regime switching models or ensemble methods are examples of future possible paths of investigation to improve the forecast. However, future work could also build upon this one. In the comparative study, all models were kept with their default parameters. Further optimization of the architecture, learning rate(s), or the number of epochs, just to name a few, would probably slightly improve forecast accuracy. The noticeable improvements resulting from the data augmentation also make us confident that including many time series datasets of other countries or related commodities traded in the year-ahead market would be valuable[4], despite being very expensive to procure.

*"An investment in knowledge pays the best interest."*

—Benjamin Franklin

# 8 Appendix

## 1 Electricity infrastructure overview

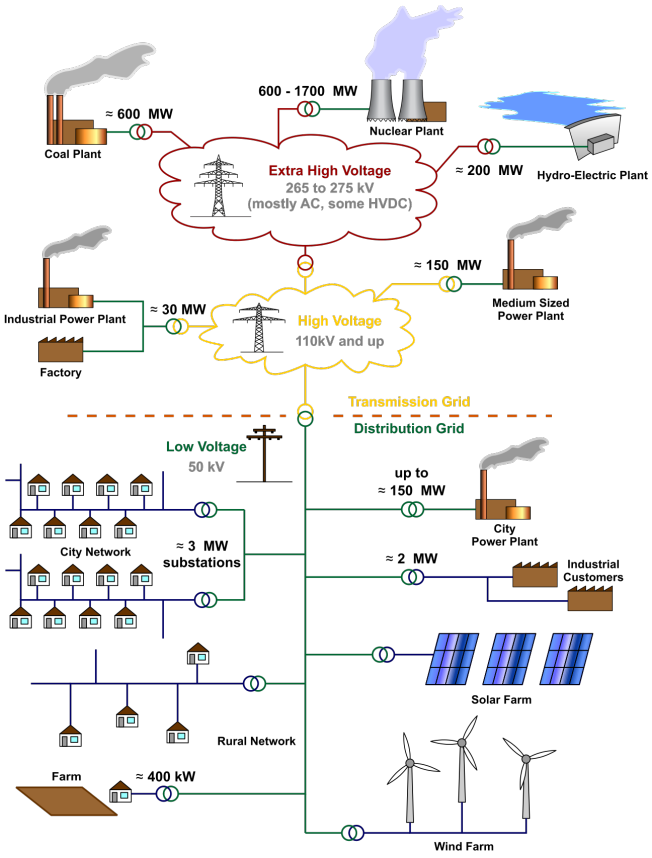



Figure 8.1 – General layout of electrical networks [8].

## 2 Electricity market overview

### 2.1 Classification of common energy generation technologies

| Type | Firm/capacity | Type of fuel | Flexibility | Low-carbon | $CO_2$ emissions (kg/KWh) |
|---|---|---|---|---|---|
| coal | firm | fossil | medium | no | 0.95 |
| natural gas | firm | fossil | high | no | 0.55 |
| biomass | firm | renewable | medium | Yes (regrowth of biomass compensates) | |
| nuclear | firm | nuclear | low | Considered as zero-emission sources | |
| hydro with dam | firm | renewable | very high | | |
| solar | variable | renewable | very low | | |
| wind | variable | renewable | very low | | |
| geothermal | firm | renewable | high | | |

Table 8.1 – Characteristics of the main energy-generation technologies [1]

### 2.1.1 The wholesale electricity market actors

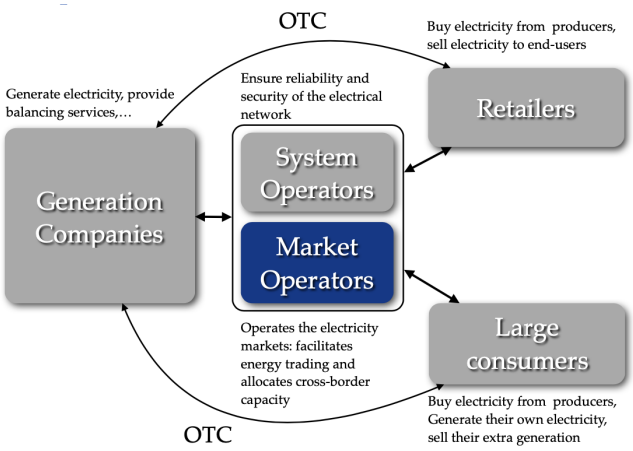



Figure 8.2 – Simplified view of market participant interactions[9].

### 2.1.2 Merit order

Figure (8.3) presents in a very simplified and schematic way the merit order effect used to define the market clearing price. Supportive legislation and the virtually free energy sources used by renewable generators can push the clearing price low enough to put other generator types out of business.

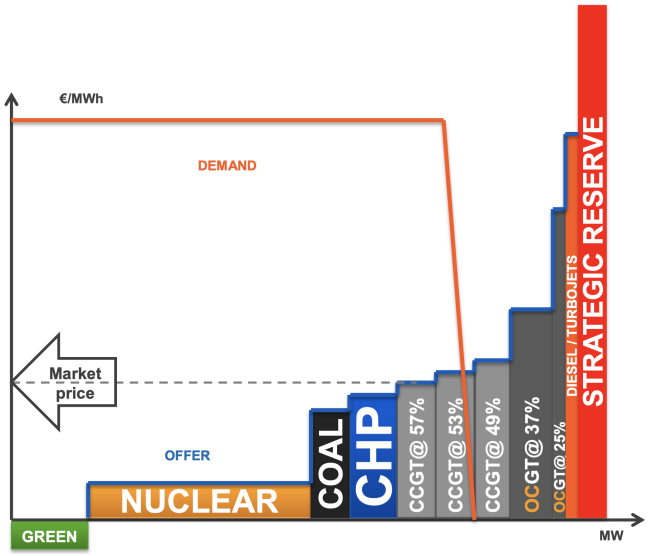



Figure 8.3 – Simplified and schematic representation of the merit order effect[3].

### 2.1.3 Market zones

As a result of the increasing market interconnection across Europe, electricity prices tend to converge among countries. Figure (8.4) shows the state of the DAM power exchange coupling as of 2015 while Figure (8.5a) reports the annual average prices for the day-ahead market from 2007 to 2018 onwards.

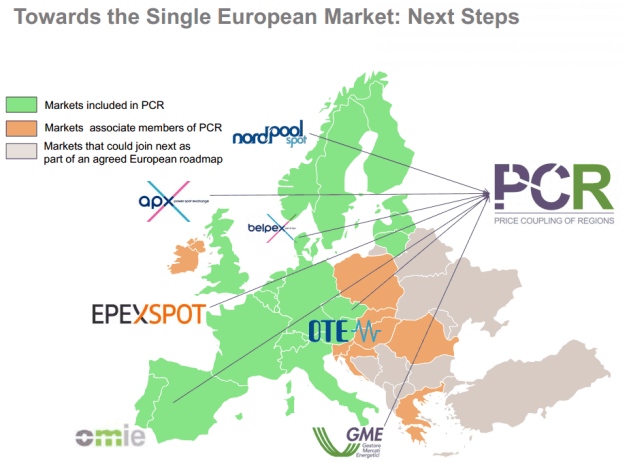
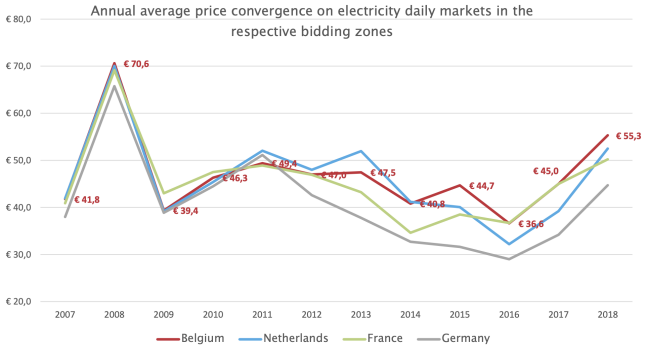
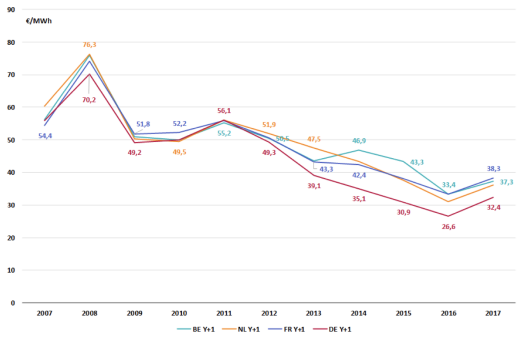



Figure 8.4 – Status of the day-ahead energy markets coupling as of 2015[10].




(a) Yearly averaged day-ahead market prices per bidding zone[23]




(b) Yearly averaged year-ahead market prices[1]

Figure 8.5 – Highlight of the electrical energy prices similarities in the CWE region.

# 3 Trading in year-ahead markets

### 3.0.1 Forward market prices comparison

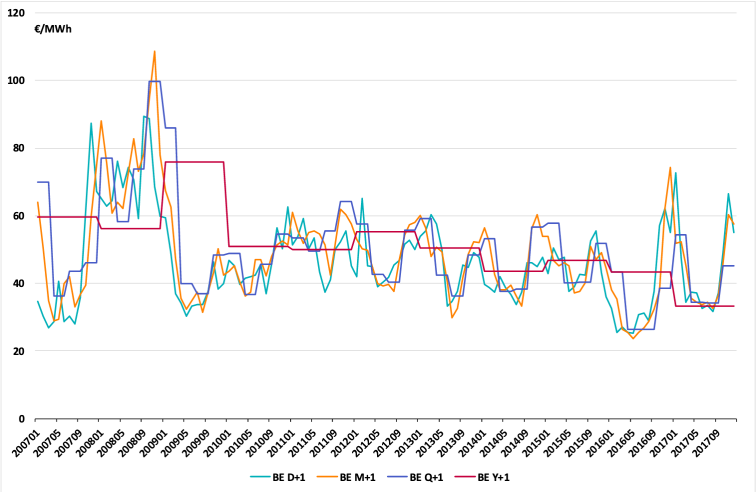



Figure 8.6 – Average prices for four types of contracts for delivery in the Belgian bidding zone, in terms of delivery period[11]. Data from ICE Endex[12] and EEX.

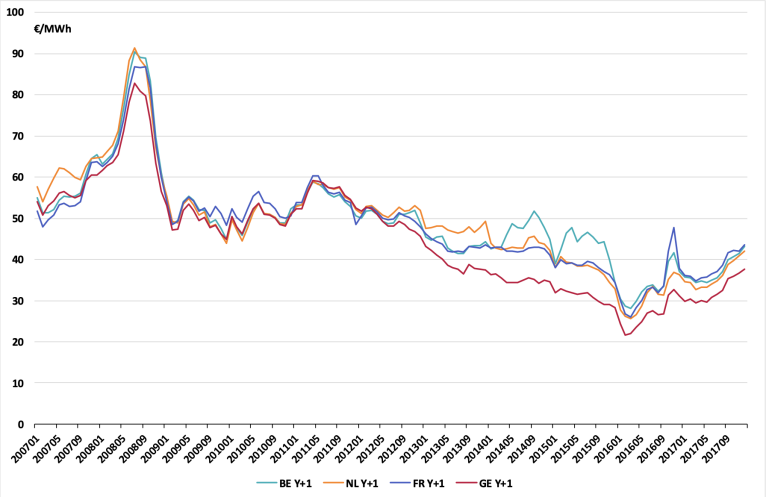



Figure 8.7 – Comparison of several wholesale power exchange market zones with monthly averaged year-ahead electricity prices[1].

# 4 Literature review

### 4.0.1 The Value-at-Risk formulation

The *Value-at-Risk*[93] is certainly one of the most widely accepted risk indicator for financial risk management. It is measured in euros and summarizes the expected maximum loss, or worst loss, of a portfolio over a given target horizon for a fixed confidence level $\alpha$. Formally, the VaR of a random variable X for $\alpha \in ]0,1[$ is:

$$VaR_\alpha(X) = \min\{z | F_X(z) \geq \alpha\} \tag{8.1}$$

Mathematically, $VaR_\alpha(X)$ is the $(1-\alpha)$ quantile of $X$. Most of the time, $\alpha$ is set to 0.95. The measure provides a straightforward interpretation to evaluate the worst case scenario after excluding all worse outcomes whose combined probability is at most $\alpha$.

# 5 Data analysis

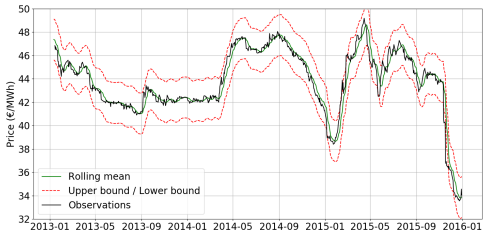
## 5.1 Smoothing techniques

**Moving average and moving median filters** Moving average and moving medians are probably the two most common types of techniques to smooth out a time series. Figure (8.8a) and Figure (8.8b) display the resulting post-processed CAL 2016 prices obtained with a window size of 10. 95% of the area under a normal curve lie within roughly 1.96 standard deviation of the mean. Due to the central limit theorem, which states that the sum of independent random variables tends towards a gaussian random variable, it is often the practice in finance to stick to a normal assumption. In econometrics, these bounds computed for a moving average of 20 gives the famous *Bollinger Bands*.

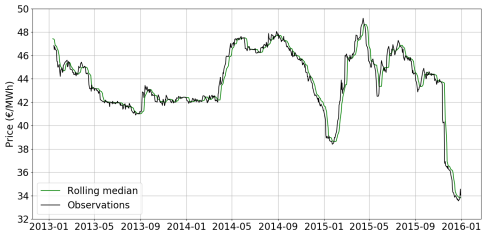
**LOESS filtering** Locally Estimated Scatterplot Smoothing (LOESS) is a generalization of moving average and polynomial regression that is popular in econometrics. Market data often expose different regimes which makes the fit of one function on the data less relevant. Instead, LOESS fit segments of the data at the cost of increased computation. In a nutshell, several low degree polynomials are fitted using weighted least squares. More weight are put on points near the point whose response is being estimated while less weight is put on points further away. This makes the method quite robust and suitable to reject sudden peaks in the data.

**EWMA filtering** Exponentially Weighted Moving Average is nothing else than an extension of moving averages in which the element within a given windows are weighted in an exponentially decaying fashion. For a same window width, EWMA is way more responsive to recent changes which make it a frequent choice in econometrics or for fault detection in quality control processes. It is computed in a self-explanatory manner according to the recursive definition:

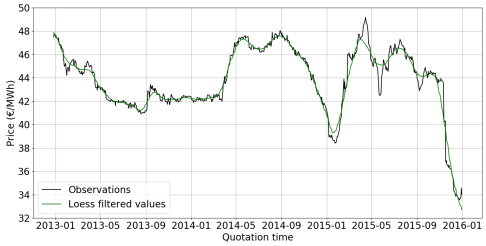
$$S_t = \begin{cases} Y_1, & t = 1 \\ \alpha \cdot Y_t + (1-\alpha) \cdot S_{t-1}, & t > 1 \end{cases} \tag{8.2}$$



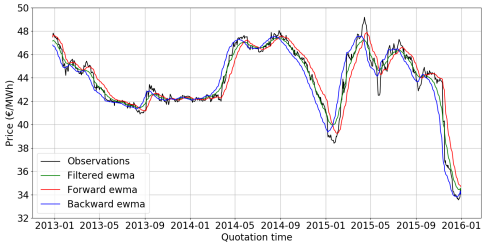

(a) Moving average filter, window size = 10




(b) Moving median filter, window size = 10




(c) LOESS filter, frac=0.05

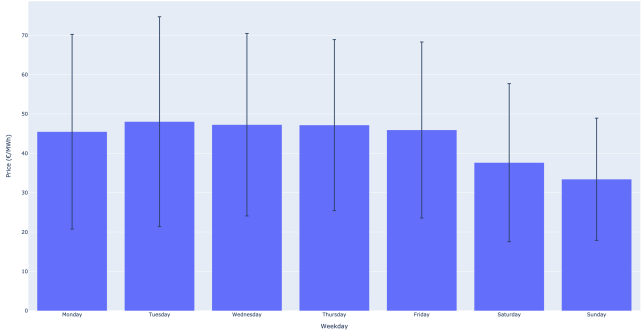



(d) EWMA filter, $\alpha = 0.095$

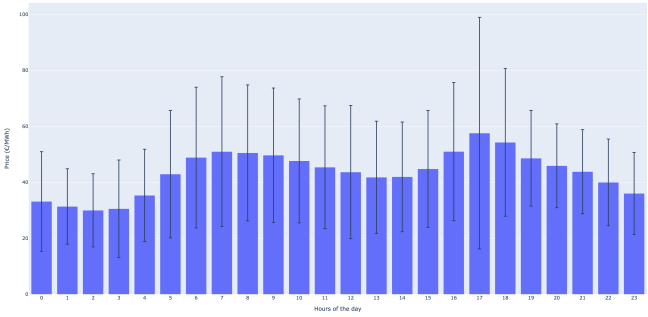
Figure 8.8 – Smoothing filter comparison

## 5.2 Pattern analysis of day-ahead market prices

Figure (8.9a) shows that the average daily prices of the DAM don't fluctuate much during working days but drop by a sensible margin during week-ends. By processing all the data from 1st May 2015 up to 22th March 2020, the average price during working days amount to 46.73€/MWh, which is 23.7% less than the regular open days of the week. The variance of the prices is also lower, especially on Sundays. The fact that many economic activities cease during the week-end cause a drop in demand of electricity which is then reflected in the price. From Figure (8.9b), we note prices during the middle of the night are, on average, way lower than in the morning and the evening. Furthermore, the variance at these time is also smaller.



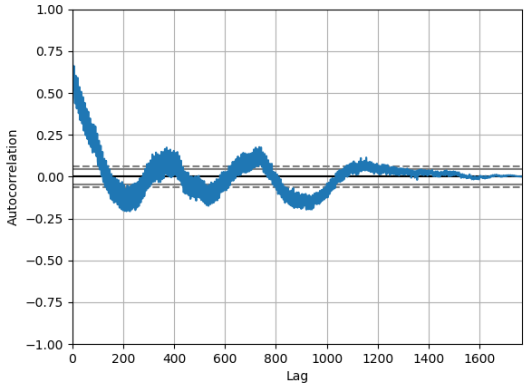

(a) Day-ahead prices per weekday



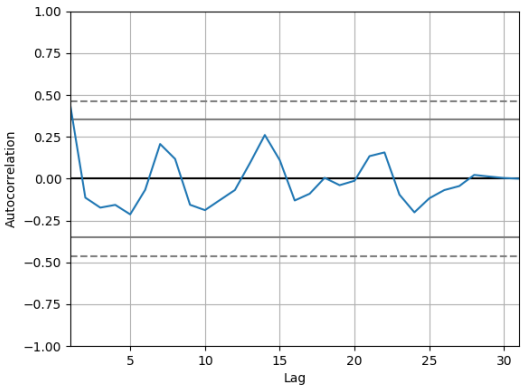

(b) Day-ahead prices per hour

Figure 8.9 – Periodicity of day-ahead prices (€/MWh)

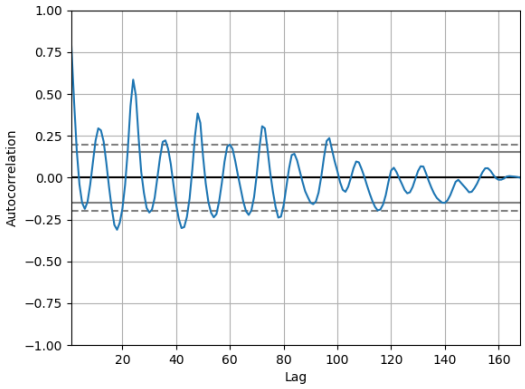
From Figure (8.10), we notice several periodic patterns rule the day-ahead prices. The weekday effect is noticeable on Figure (8.10b) while Figure (8.10c) shows a periodic pattern every 24h hours (larger spikes), and a smaller one of around 15 hours approximately. In this way, despite being very volatile, day-ahead prices are not are not of fully stochastic nature.




(a) All DAM dataset. [lag] = 1 day.




(b) March 2016 DAM prices, [lag] = 1 day.

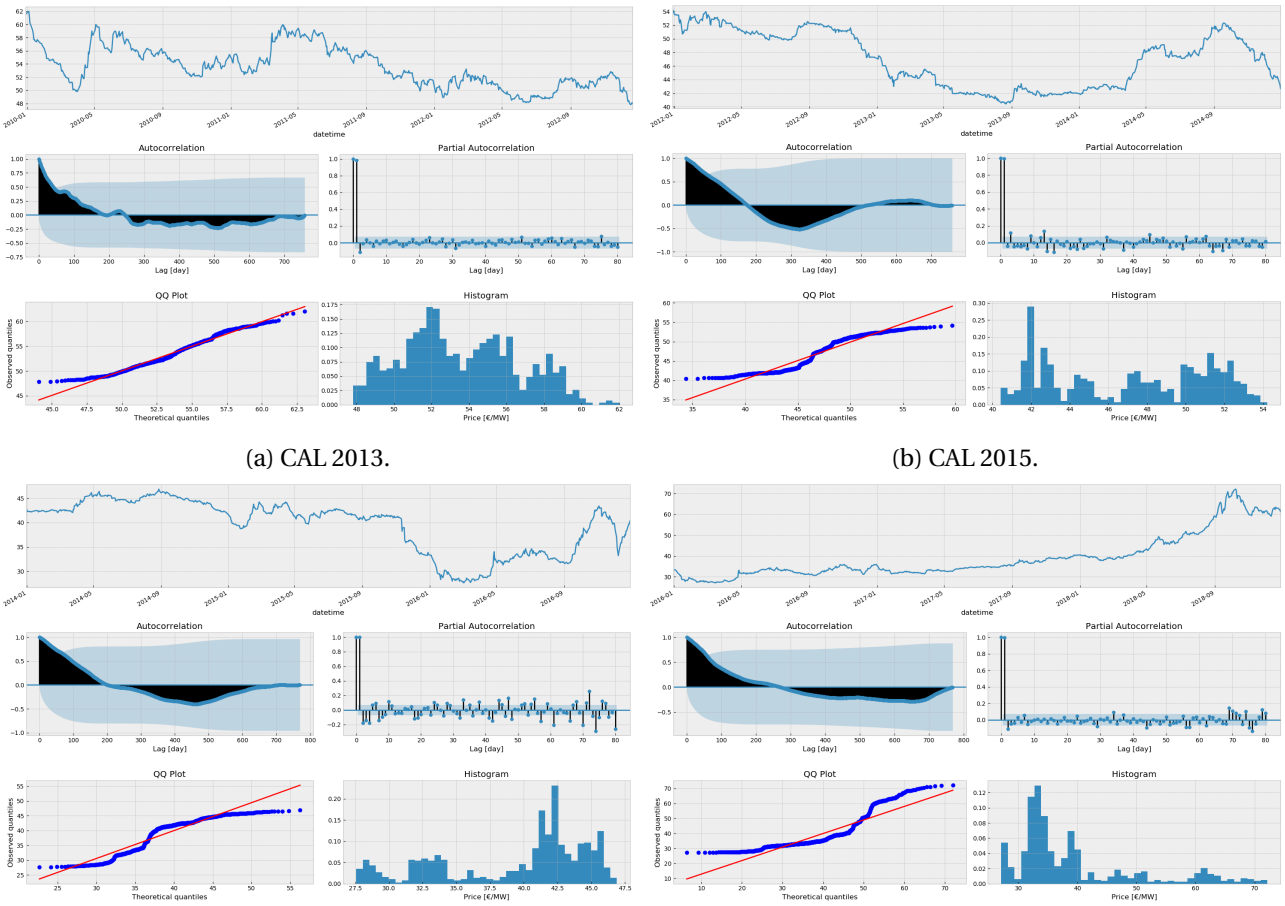



(c) Week 10 of 2016 (in March) DAM prices, [lag] = 1 hour.

Figure 8.10 – Autocorrelation of DAM prices at different time scales and/or time resolution. Data is aggregated based on the median.

## 5.3 Supplement on year-ahead market prices analysis

### 5.3.1 Summary plots for four additional CAL




(a) CAL 2013.

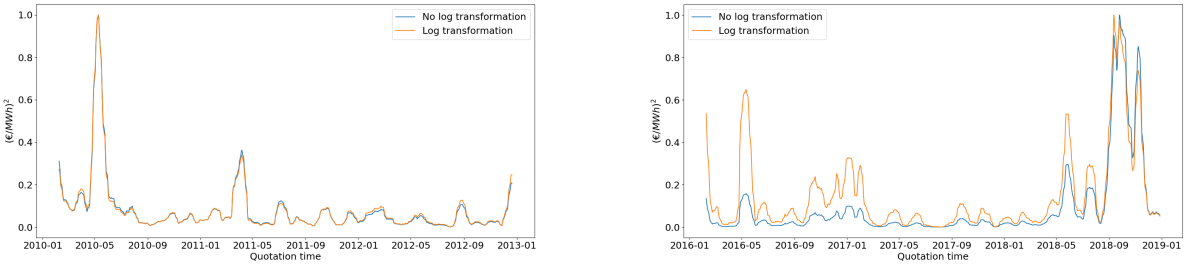
(b) CAL 2015.

(c) CAL 2017.

(d) CAL 2019.

Figure 8.11 – Data analysis summary plots for four different CAL of the year-ahead market.

### 5.3.2 Impact of the log transformation on CAL 2013 and CAL 2019




(a) CAL 2013

(b) CAL 2019

Figure 8.12 – Rolling variance of year-ahead market prices with a window size of 28 days (four weeks).

## 5.4 Supplement on exogenous variable discussion

### 5.4.1 EU allowances

EU Allowances (EUAs) are a form of carbon allowance used as the main currency in the EU Emissions Trading Scheme (EU ETS). Companies are allowed to trade those carbon credits between themselves. The majority of EUA carbon trading takes place on exchanges, with the Intercontinental Exchange (ICE), being the most prominent one. The ECX EU Futures (Continuous Contract #1 from ICE), fetched from quandl, can

provide additional meaningful information to the improve forecast accuracy of the YAM prices. One EUA gives the holder the right to emit one tonne of carbon dioxide, or the equivalent amount of two more powerful greenhouse gases, nitrous oxide (N2O) and perfluorocarbons (PFCs). The unit is Euro (€) per metric tonne and the data has a weekly frequency.

Figure (8.13) displays all the scaled CAL prices in low intensity orange with the mean of concurrently quoted cal in deep orange for easier patterns interpretation. The purple trace represents the scaled carbon prices, fetched from the Global European Open platform. Leader-follower roles seem to change overtime.


Figure 8.13 – Carbon prices and all year ahead CAL along with their average.

### 5.4.2 All exogenous variables

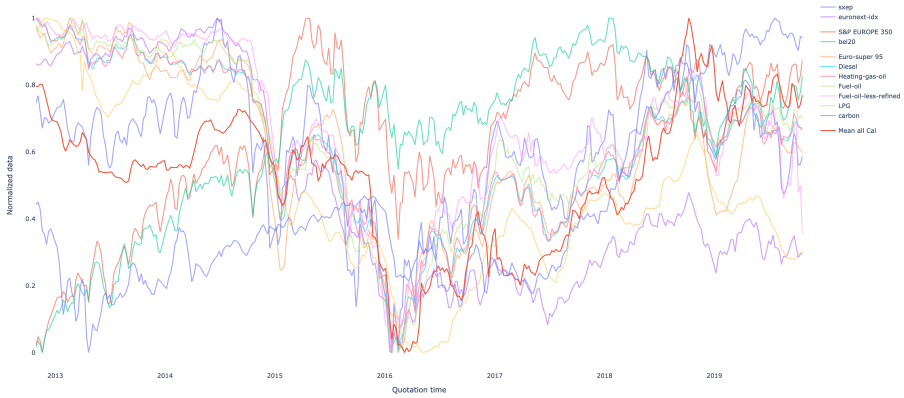



Figure 8.14 – All exogenous variables alongside the Mean of all cal. Data aggregated weekly, logged transformed then scaled in [0, 1].

## 6 Forecasting

### 6.0.1 Handling the case of concurrently quoted CAL

Let denote by $t_0$ the forecast start date. To handle the inclusion of concurrently quoted CAL data in the training set at $t_0$, we only include CAL whose quotation period can be mapped one to one with the CAL being forecasted. Figure (8.15) illustrate this mechanism for CAL 2016 for three scenarios. When $t_0$ is located in the first year of CAL 2016, i.e. *scenario A*, CAL 2014 and CAL 2015 are included in the training set. In *scenario B*, only CAL 2015 is added to the training set. Finally, in *scenario C*, no other CAL has shared the exact same number of open days as CAL 2016. For this reason, no other concurrently quoted CAL are used in the training set for the corresponding $t_0$. While not being the most data efficient mechanism, it certainly removes any possible miss-guidance due to zero-padding. Many different alternatives can of course

be derived. Introducing categorical features or using a cold start approach (see MQRNN[6] to improve the forecast accuracy in the very beginning of the quotation period are natural extension that can be investigated.

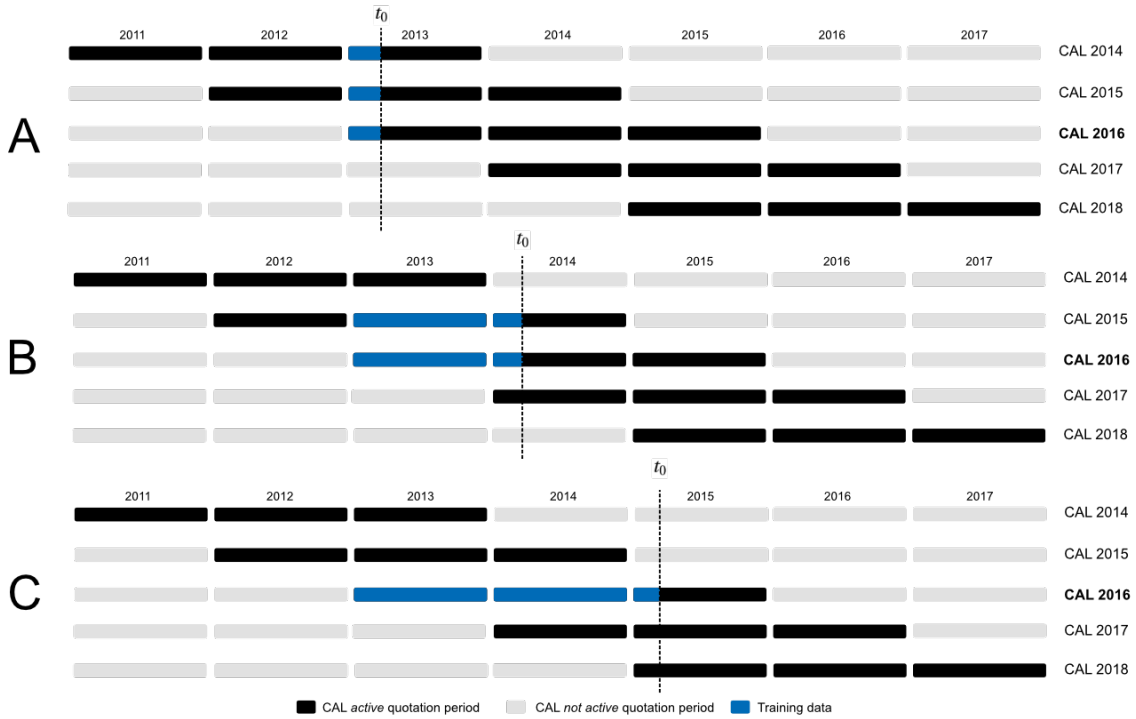



Figure 8.15 – Illustration of the inclusion criterion for concurrently quoted CAL for three forecast start time ($t_0$) scenario, assuming CAL 2016 is the CAL of interest.

### 6.0.2 Normality assessment of residuals.

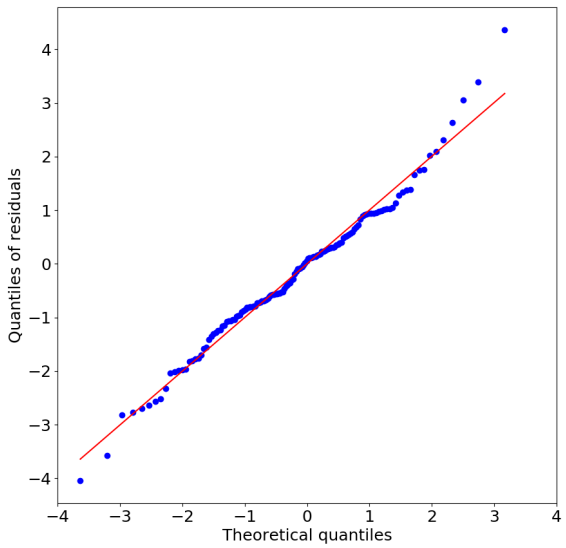



Figure 8.16 – QQ plot of the forecast residuals produced by the backtesting simulation of Simple Feed Forward Neural Network on CAL 2013.

# Bibliography


[1] Gregor Erbach (European Parliamentary Research Service). Understanding electricity marketsin the eu. https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593519/EPRS_BRI(2016)593519_EN.pdf, November 2016.

[2] Deloitte. European energy market reform, country profile: Belgium. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Energy-and-Resources/gx-er-market-reform-belgium.pdf. [Online; accessed 4-March 2020].

[3] Frédérik Demaret (B2B Services EDF Luminus). Elec0018-1 (uliège) - energy trading.

[4] David Salinas, Valentin Flunkert, and Jan Gasthaus. Deepar: Probabilistic forecasting with autoregressive recurrent networks, 2017.

[5] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio, 2016.

[6] Ruofeng Wen, Kari Torkkola, Balakrishnan Narayanaswamy, and Dhruv Madeka. A multi-horizon quantile recurrent forecaster, 2017.

[7] Courtney Cochrane. Time series nested cross-validation. https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9. Accessed 2020-05-26.

[8] Electrical grid. https://en.wikipedia.org/wiki/Electrical_grid.

[9] Bertrand Cornélusse. Elec0018-1 (uliège) - how the eu day-ahead electricity market works (slides). http://blogs.ulg.ac.be/damien-ernst/teaching/elec0018-1-energy-markets/.

[10] Epex Spot. Price coupling in europe. http://www.epexspot.com/en/market-coupling/pcr.

[11] CREG Commission de l'électricité et du gaz. Study on the functioning and price evolution of the belgian wholesale electricity market – monitoring report 2017.

[12] Ice endex. https://www.theice.com/endex.

[13] Boris Defourny, Damien Ernst, and Louis Wehenkel. Multistage stochastic programming: A scenario tree based approach to planning under uncertainty. *LE, Sucar, EF, Morales, and J., Hoey (Eds.), Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions. Hershey, Pennsylvania, USA: Information Science Publishing*, 01 2011.

[14] Swde. https://www.swde.be/fr.

[15] SWDE. Official 2018 report. https://www.swde.be/en/media-library/documentation/annual-report.

[16] Engie electrabel. https://www.engie.be/fr/.

[17] Sébastien Mathieu. *Flexibility services in the electrical system.* PhD thesis, Université de Liège, Liège, Belgique, 2016.

[18] KU Leuven Energy Institute. The current electricity market design in europe. https://set.kuleuven.be/ei/images/EI_factsheet8_eng.pdf/, 2015.

[19] Elia. https://www.elia.be/.

[20] Synergrid. http://www.synergrid.be/index.cfm?PageID=16823#.

[21] Léonard Wagner. Chapter 27 - overview of energy storage technologies. In Trevor M. Letcher, editor, *Future Energy (Second Edition)*, pages 613 – 631. Elsevier, Boston, second edition edition, 2014.

[22] Alex Benjamin WILSON (European Parliamentary Research Service). Smart electricity grids and meters in the eu member states. https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2015)568318, September 2015.

[23] Febeg. Annual report 2018. https://www.febeg.be/sites/default/files/febeg_annual_report_slide_deck_2018_fr_final.pdf.

[24] Elia press release 2018. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=2ahUKEwjY8sLMueLnAhXP5KQKHTp8Cp4QFjACegQIAhAB&url=https%3A%2F%2Fwww.elia.be%2F-%2Fmedia%2Fproject%2Felia%2Fshared%2Fdocuments%2Fpress-releases%2F2019%2F20190118_energymix%2F20190118_press-release-chiffres-2018_en.pdf&usg=AOvVaw0aJm19tCJDkJtWoNl5tWiE.

[25] Electricity-map. https://www.electricitymap.org/?page=country&countryCode=BE.

[26] EUROPEAN COMMISSION). State of the energy union 2015 - communication from the commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1449767367230&uri=CELEX:52015DC0572, November 2015.

[27] Network codes. https://ec.europa.eu/energy/en/topics/markets-and-consumers/wholesale-energy-market/electricity-network-codes.

[28] Damien Ernst. Elec0018-1 (uliège) - energy market lecture slides. http://blogs.ulg.ac.be/damien-ernst/teaching/elec0018-1-energy-markets/.

[29] Electricity and time factors. https://www.next-kraftwerke.be/fr/plateforme-de-connaissances/les-marches-de-lelectricite-et-le-facteur-temps/.

[30] Entso-e market report 2019. https://todo.com.

[31] Lion Hirth, Jonathan Mühlenpfordt, and Marisa Bulkeley. The entso-e transparency platform – a review of europe's most ambitious electricity data platform. *Applied Energy*, 225:1054 – 1067, 2018.

[32] Koen Rademaekers et al. (European Commission). Study on energy prices, costs and subsidies and their impact on industry and households. 2017.

[33] Aleasoft. https://aleasoft.com/fr/.

[34] N-side. https://www.n-side.com/solution/electricity-price-forecasts/.

[35] Elexys. https://my.elexys.be/MarketInformation/IceEndexPowerBE.aspx.

[36] Paula Rocha and Daniel Kuhn. Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules. *European Journal of Operational Research*, 216(2):397 – 408, 2012.

[37] Raphaël Homayoun Boroumand, Stéphane Goutte, Simon Porcher, and Thomas Porcher. Hedging strategies in energy markets: The case of electricity retailers. *Energy Economics*, 51, 07 2015.

[38] Min Liu and Felix Wu. Portfolio optimization in electricity markets. *Electric Power Systems Research*, 77:1000–1009, 06 2007.

[39] Sheena Yau, Roy H. Kwon, J. Scott Rogers, and Desheng Wu. Financial and operational decisions in the electricity sector: Contract portfolio optimization with the conditional value-at-risk criterion. *International Journal of Production Economics*, 134(1):67 – 77, 2011. Enterprise risk management in operations.

[40] Ákos Baldauf. Trading techniques for european electricity markets. 04 2018.

[41] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30, 10 2014.

[42] Gergo Barta, Gyula Borbely Gabor Nagy, Sandor Kazi, and Tamas Henk. Gefcom 2014—probabilistic electricity price forecasting. *Smart Innovation, Systems and Technologies*, page 67–76, 2015.

[43] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob Hyndman. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, 32, 03 2016.

[44] Asset definition.

[45] Jakub Nowotarski and Rafał Weron. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81, 06 2017.

[46] Athanasopoulos G. Hyndman, R.J. Forecasting: principles and practice. `https://otexts.com/fpp2/`, 2018. 2nd edition.

[47] Spyros Makridakis, Evangelos Spiliotis, and Vassilis Assimakopoulos. The m4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 06 2018.

[48] Jai Ranganathan Slawek Smyl and Andrea Pasqua. M4 forecasting competition: Introducing a new hybrid es-rnn model. `https://eng.uber.com/m4-forecasting-competition/`.

[49] Spyros Makridakis, Evangelos Spiliotis, and Vassilis Assimakopoulos. M4 competitor's guide: Prizes and rules.

[50] Jurgen A. Doornik Jennifer L. Castle and David F. Hendry. Some forecasting principles from the m4 competition. January 2019.

[51] M4 data repository. `https://mofc.unic.ac.cy/the-dataset/`. https://github.com/Mcompetitions/M4-methods.

[52] Peter Zhang, Eddy Patuwo, and Michael Hu. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14:35–62, 03 1998.

[53] A. Alexandrov, K. Benidis, M. Bohlke-Schneider, V. Flunkert, J. Gasthaus, T. Januschowski, D. C. Maddix, S. Rangapuram, D. Salinas, J. Schulz, L. Stella, A. C. Türkmen, and Y. Wang. GluonTS: Probabilistic Time Series Modeling in Python. *arXiv preprint arXiv:1906.05264*, 2019.

[54] Nikolay Laptev, Jason Yosinski, Li Erran Li, and Slawek Smyl. Time-series extreme event forecasting with neural networks at uber. 2017.

[55] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179 – 211, 1990.

[56] Felix Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. *Neural computation*, 12:2451–71, 10 2000.

[57] Alex Graves. Generating sequences with recurrent neural networks, 2013.

[58] Rob J Hyndman and Yeasmin Khandakar. Automatic time series forecasting: the forecast package for r, 2008.

[59] George E. P. Box and Gwilym M. Jenkins. *Some recent advances in forecasting and control.* 1968.

[60] Benjamin Letham Sean J Taylor. Forecasting at scale. *PeerJ Preprints*, Septembre 2017.

[61] Amazon Web Service (AWS). Time series forecasting principles with amazon forecast: Technical guide. https://d1.awsstatic.com/whitepapers/time-series-forecasting-principles-amazon-forecast.pdf?did=wp_card&trk=wp_card, February 2020.

[62] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. Deep state space models for time series forecasting. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 7785–7794. Curran Associates, Inc., 2018.

[63] Guillaume Chevillon. Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21:746 – 785, 09 2007.

[64] Wikipedia contributors. Partial autocorrelation function — Wikipedia, the free encyclopedia, 2019. [Online; accessed 20-March-2020].

[65] M. D. Scheuerell E. E. Holmes and E. J. Ward. *Applied Time Series Analysis for Fisheries and Environmental Sciences*. NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd E., Seattle, WA 98112.

[66] Selva Prabhakaran. Augmented dickey fuller test (adf test) – must read guide. https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/.

[67] Clifford M. Hurvich. Differencing and unit root tests. http://pages.stern.nyu.edu/~churvich/Forecasting/Handouts/UnitRoot.pdf.

[68] Joe Rodgers and Alan Nicewander. Thirteen ways to look at the correlation coefficient. *American Statistician - AMER STATIST*, 42:59–66, 02 1988.

[69] C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904.

[70] Jin Hyun Cheong. Four ways to quantify synchrony between time series data. https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9.

[71] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[72] Michael Eichler. Causal inference in time series analysis. http://researchers-sbe.unimaas.nl/michaeleichler/wp-content/uploads/sites/31/2014/02/causalstatistics.pdf.

[73] Eua futures from ice. https://www.quandl.com/data/CHRIS/ICE_C1-ECX-EUA-Futures-Continuous-Contract.

[74] Europa oil bulletin data. https://data.europa.eu/euodp/en/data/dataset/eu-oil-bulletin/resource/24897252-8a68-4ac4-80f2-7c9360c04d14.

[75] S&p 350 stock index. https://us.spindices.com/indices/equity/sp-europe-350.

[76] Bel 20 stock index. https://fr.investing.com/indices/bel-20-historical-data.

[77] Consumer price index data. https://data.oecd.org/fr/price/inflation-ipc.htm#indicator-chart.

[78] Stoxx europe 600: oil and gas sxep index. https://fr.investing.com/indices/stoxx-europe-600-oil---gas-historical-data.

[79] Ricie euronext rogers international energy commodity index. https://www.quandl.com/data/RICI/RICIE-Euronext-Rogers-International-Energy-Commodity-Index.

[80] Helmut Luetkepohl and Fang Xu. The role of the log transformation in forecasting economic variables. *Empirical Economics*, 42, 03 2009.

[81] Nagesh Singh Chauhan. Stock market forecasting using time series. https://towardsdatascience.com/stock-market-forecasting-using-time-series-c3d21f2dd37f.

[82] Sarit Maitra. Time series forecasting using granger's causality and vector auto-regressive model. https://towardsdatascience.com/granger-causality-and-vector-auto-regressive-model-for-time-series-forecasting-3226a64889a6.

[83] Yann Lecun, Leon Bottou, Genevieve Orr, and Klaus-Robert Müller. Efficient backprop. 08 2000.

[84] Frederic Stallaert. Benchmarking gaussian processes for time series forecasting. Master's thesis, Gent University, 2017.

[85] Yuyang Wang, Alex Smola, Danielle C. Maddix, Jan Gasthaus, Dean Foster, and Tim Januschowski. Deep factors for forecasting, 2019.

[86] Leonard J. Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437 – 450, 2000. The M3- Competition.

[87] Christoph Bergmeir and José Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 05 2012.

[88] Rob Hyndman and Anne Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688, 02 2006.

[89] J.S. Armonstrong. Long-range forecasting: From cystal ball to computer, 1978.

[90] A.E. Raftery T. Gneiting. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, pages 359–378, 2007.

[91] Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

[92] Google colaboratory platform. https://colab.research.google.com/notebooks/basic_features_overview.ipynb.

[93] Sergey Sarykalin, Gaia Serraino, and Stan Uryasev. Value- at-risk vs conditional value-at-risk in risk management and optimization. 09 2008.