Universität Augsburg

Institut für Mathematik

Continuous-Time Autoregressive Processes for Modeling Electricity Prices and Renewable Energies

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
eingereicht an der
Mathematisch-Naturwissenschaftlich-Technischen Fakultät
der Universität Augsburg

von

Daniel Lingohr

2019

Erstgutachter:	Prof. Dr. Gernot Müller
Zweitgutachter:	Prof. Dr. Ralf Werner
Drittgutachter:	Prof. Dr. Ralf Korn
	(Technische Universität Kaiserslautern)

Tag der mündlichen Prüfung: 24.06.2019

Acknowledgement

First of all I would like to thank my supervisor Prof. Dr. Gernot Müller, who always supported me and contributed to the realization of my ideas trough many fruitful discussion. His professional feedback helped significantly to improve the quality of this thesis. I would also like to thank him for always providing a very comfortable working atmosphere at the chair.

Of course, my other colleagues also had a part in this, and I would also like to thank them for the great time.

Further I particularly would like to thank Jean Jacod for his constructive comments on the contents of Chapter 4 and the discussions on the proofs given there.

Last but not least I would like to thank my father, who always backed me up in my career and made it possible that I could acquire the higher education entrance qualification and complete my studies after my apprenticeship.

Abstract

This thesis deals with stochastic models for electricity markets. The focus is on wholesale prices and renewable power generation. Continuous-time autoregressive (CAR) processes are frequently used in this context. We define two new types within the class of CAR processes and show their benefits in application to the electricity market. The first process extends CAR processes with regime switching mean-reversion rates by including a jump component. This is necessary because spikes are one of the most pronounced features of electricity prices. CAR processes with non-constant parameters are also considered in an innovative model for photovoltaic (PV) power generation. This model provides for the first time a pure statistical approach to map intraday variation of solar power infeed. The second newly defined stochastic process allows to include external information in a flexible way. This makes it possible to take many facets of renewable energy production into account when determining the electricity price. Since renewable energies have an increasing impact on electricity prices, models that can handle related information are becoming more and more important. All results are applied to the German electricity market. Implementations for the newly defined processes are provided in R and C++.

Zusammenfassung

Inhalt dieser Arbeit sind stochastische Verfahren zur Modellierung von Strommärkten. Der Schwerpunkt liegt dabei auf der Beschreibung von Großhandelspreisen und erneuerbarer Stromerzeugung. Zeitstetige autoregressive Prozesse werden in diesem Zusammenhang häufig eingesetzt. Wir definieren zwei neue Typen solcher Prozesse und zeigen welche Vorteile sich durch ihre Anwendung ergeben. Der erste Prozess erweitert autoregressive Zeitreihen mit zustandsabhängigen Rückkehrraten durch Hinzufügung einer Sprungkomponente. Dies ist notwendig, da Preisspitzen eine wesentliche Chrakteristik von Strompreisen darstellen. Autoregressive Prozesse mit nicht konstanten Parametern werden auch in einem innovativen Modell zur Beschreibung der Stromerzeugung von Photovoltaikanlagen berücksichtigt. Dieses Modell ermöglicht erstmals die Abbildung der untertägigen Variation des produzierten Solarstroms mittels eines rein statistischen Ansatzes. Der zweite neu definierte stochastische Prozess ermöglicht es, externe Informationen flexibel einzubinden. Dies wird genutzt um viele Facetten der erneuerbarer Energiegewinnung bei der Ermittlung des Strompreises zu berücksichtigen. Da erneuerbare Energien einen zunehmenden Einfluss auf die Strompreise haben, gewinnen Modelle, die mit entsprechenden Informationen umgehen können, zunehmend an Bedeutung. Unsere Ergebnisse werden in ihrer Anwendung für den deutschen Strommarkt untersucht. Implementierungen, die die neu definierten Prozesse betreffen, werden in R und C++ bereitgestellt.

Contents

Chapter 1

Introduction

1.1 Context

Electricity is a commodity with specific features that distinguish it from other energy sources. Electricity is immaterial, homogeneous and at the time of writing has still very limited storage possibilities (Graeber, 2014). Supply and demand must therefore always match. The electricity market thus differs substantially from other commodity markets. This has led to a need for new types of models on which market participants can base their decisions. Power portfolio managers rely on specific electricity price models to schedule production or consumption with the aim of maximizing profit. Such models are also essential to value derivatives, which are required to counter price risks. The strong increase in renewable power generation has made these risks even greater and added significant volume risks. Accurate modeling of the electricity price as well as the impact of renewable energy sources has therefore become an important task.

Due to the lack of storability, electricity has unique price dynamics. A strong seasonality and high volatility are part of its characteristics. Jumps occur frequently but have only a short impact on the price. Reversion to the mean is a phenomenon that can be observed not only for electricity prices but also for related natural processes. This includes renewable power generation. Autoregressive processes model such dynamics in a natural way. Therefore, they are the basis of many electricity price models and also our research is build on them.

Electricity is generally referred to as the flow of electric charge. To move the charge against the resistance of the transmission grid, power plants must generate voltage. The charge flowing through the grid provides electrical power. This power is then used over a specific period of time. The electricity price is therefore the price of electrical power

delivered over a certain time interval. It thus makes sense to consider a continuous-time autoregressive process representing the price for an infinitesimal delivery time (cf. Benth et al. (2008)). This is in line with financial mathematics in which continuous-time processes in general have become increasingly important due to their successful application in the Black-Scholes model. Since electrical power is continuously provided, it is also natural to use continuous-time processes to model renewable power generation. This also has the advantage that it is easy to deal with weather-related power plant downtimes.

Based on the above, in this thesis we consider stochastic models for electricity markets which rely on continuous-time autoregressive processes. In detail, we build models for intraday market prices as well as solar power generation. Also a new approach for modeling daily average prices is proposed. The academic contribution of this work is twofold. First, new stochastic methods for analyzing time series are introduced. These could be useful in a wide range of applications. Second, the theoretical models are adapted to the electricity market and their benefits are demonstrated. This is to contribute to energy economics.

The main results are

- an extension of nonlinear autoregressive processes such that jumps are possible,
- a model for photovoltaic (PV) power generation that takes intraday variations into account, and
- the introduction of a stochastic process to map information on renewable power generation to electricity prices in a flexible way.

All methods studied are checked for application to the German electricity market, including calibration, validation and comparison to existing approaches. For this, we use various data and implement the new techniques using the software environment R. We also use parallelized C++ code for a numerically complex estimation procedure.

The main parts of Chapter 3 will appear in Lingohr and Müller (2019). Contents of Chapter 2 and Chapter 4 are submitted for publication.

1.2 Structure of the thesis

The remaining part of Chapter 1 is intended to provide a deeper insight into the motivation behind this thesis. To this end, we give an overview of the German electricity market and in particular address the importance of renewable energies. We discuss their

impact on the electricity price and review existing approaches to model the dependence between price and renewable energy sources. We also summarize some basic results on continuous-time autoregressive processes.

In Chapter 2 we consider continuous-time threshold autoregressions (CTAR) as introduced by Brockwell and Hyndman (1992). CTAR are nonlinear autoregressive processes that exhibit a regime switching behavior. So far, only Gaussian CTAR have been defined. This leads to a lack in modeling data with jumps, as frequently observed in the electricity market and other financial markets as well. We fill this gap by constructing a CTAR with jumps. Existence of a unique weak solution and weak consistency of an Euler approximation scheme is proven. As a closed form expression of the likelihood is not available, we use kernel based particle filtering for estimation. We apply the model to the physical electricity index (Phelix) at the European Energy Exchange (EEX) and show its benefits compared to other approaches.

Chapter 3 analyzes the time series of PV power generation. The fact that no solar power is generated at night makes modeling of this times series for high resolution difficult. Previous work has therefore been limited to daily variations. In contrast, we propose a model that is able to take intraday effects into account. To do this, first we filter a cloud cover component from the infeed data by using physical relationships. This variable incorporates the complete stochastics and can be modeled by CTAR as treated in Chapter 2. We fit our model to infeed data in Germany and show its benefits compared to other approaches. The model enables pricing of derivatives, which is illustrated by a new future contract. This product allows the volume risk of PV power plants to be hedged.

Chapter 4 is devoted to the question of how the effects of renewable energies on electricity prices can be modeled. We offer a new approach that makes it possible to represent this dependency in a flexible way. It will especially be possible to vary price spikes both in frequency and size according to the generation of wind and solar power. The proposed concept is based on the construction of a process with conditionally independent increment (CII). We provide an estimation procedure for this class of processes and offer a test to verify the dependency on an external variable. Based on the theoretical results, we fit an electricity price model for the German intraday market and use it to analyze power cap/floor futures.

The results of this thesis are summarized in Chapter 5. There we also give an outlook on aspects arising from our work that could serve as an incentive for further research.

Appendix A contains the essential R and C++ source code of our implementations. This includes functions for computing a likelihood based on particle filtering as dealt with in Chapter 2, calculating an object's solar angle as discussed in Chapter 3 and the nonparametric estimator and the test introduced in Chapter 4.

1.3 The German electricity market

Electricity markets have undergone major changes in recent decades. In addition to structural and organizational changes brought by liberalization, power generation itself changed. Renewable energies are playing an increasingly important role. The following is an overview of the history and present structure of the German electricity market. As a comprehensive study on this topic which also deals more with technical details, Graeber (2014) is recommended.

1.3.1 The liberalized market

With the passage of the Gesetz über die Elektrizität- und Gasversorgung (Law on Energy Management) (EnWG) in April 1998 the basis for the today's German power market was set. This law should ensure that the future supply of electricity is as safe, cheap and environmentally as friendly as possible. To achieve the goals of the EnWG a competitive electricity market was established. The monopolies of the energy suppliers were ended, giving end-users the opportunity to freely choose their retailers. Moreover, the electricity companies were obliged to make their transmission networks available to others as well. For this to take place under fair conditions, production, transport and distribution had to be unbundled. This ultimately led to the present form of the German electricity market, consisting of pure grid operators in charge of transport and distribution as well as companies responsible for generation, trading and sales.

1.3.2 Supply and demand

With an net electricity consumption of 513 TWh in 2017 (BNetzA (2018)), the German electricity market is the largest power market in Europe. The industry is by far the biggest consumer with a share of almost a half. The remaining consumption comes almost equally from households and trade and commerce. Demand by traffic only plays a minor role (cf. Table 1.1). Historical demand in terms of the total load in Germany is shown in Figure 1.1. We observe large variations depending on season and time of day. In particular, there is lower demand at summer and night. Holiday seasons and weekends

Table 1.1: Net electricity consumption in 2017 by consumer groups (AGEB (2017)).

Industry	Trade and Commerce	Households	Traffic
47%	26%	25%	2%

also have a reducing effect. This periodicity is mainly the result of different business activities and climatic conditions. Thus, the electricity load in winter increases due to a higher need in lighting and heating. Considering the intraday profiles, also a "double peak structure" is displayed. In the afternoon the general business activity decreases, while private activities at the end of the working day trigger an additional demand. Because of longer daytime, this double peak structure is less pronounced in summer.

Figure 1.2 shows how demand is covered in Germany. Lignite, as a cheap fossil fuel, is used for the most part. The share of electricity from nuclear power plants has been reduced significantly in recent decades – 13 % in 2017 compared to 31 % in 2002. As a result of the Fukushima nuclear catastrophe in 2011, it was politically decided to end nuclear power generation in Germany by 2022. Gas-fired power plants now provide the largest capacity of all conventional energy sources. As a flexible technology, they are particularly important as a supplement to variable production from renewable energies. Hence, they play a major role for the energy transition (Energiewende) in Germany.

1.3.3 Renewable energies

Increasing global demand for energy requires an expansion of power generation. The finite nature of fossil fuels and the large quantities of carbon dioxide they produce have pushed the search for alternative energy sources. The need for renewable energies to cover energy consumption and reduce greenhouse gases is now recognized worldwide. Thus, almost all countries of the world support renewable energy development now. At the 2016 United Nations Climate Change Conference, the leaders of 48 developing countries even committed themselves to working towards a complete supply by renewable energies. This includes also Germany.

Figure 1.3 shows how energy sources changed between 2000 and 2017. Electricity generation from conventional power plants was constantly reduced and replaced by renewable energies. By 2017, 33 % of the total gross electricity generation came from renewable sources. This is an increase of 26 % compared to 2000. The basis for this development was the Gesetz für den Ausbau erneuerbarer Energien (Renewable Energy Sources Act) (EEG) passed in 2000. In interest of climate and environmental protection, it was decided

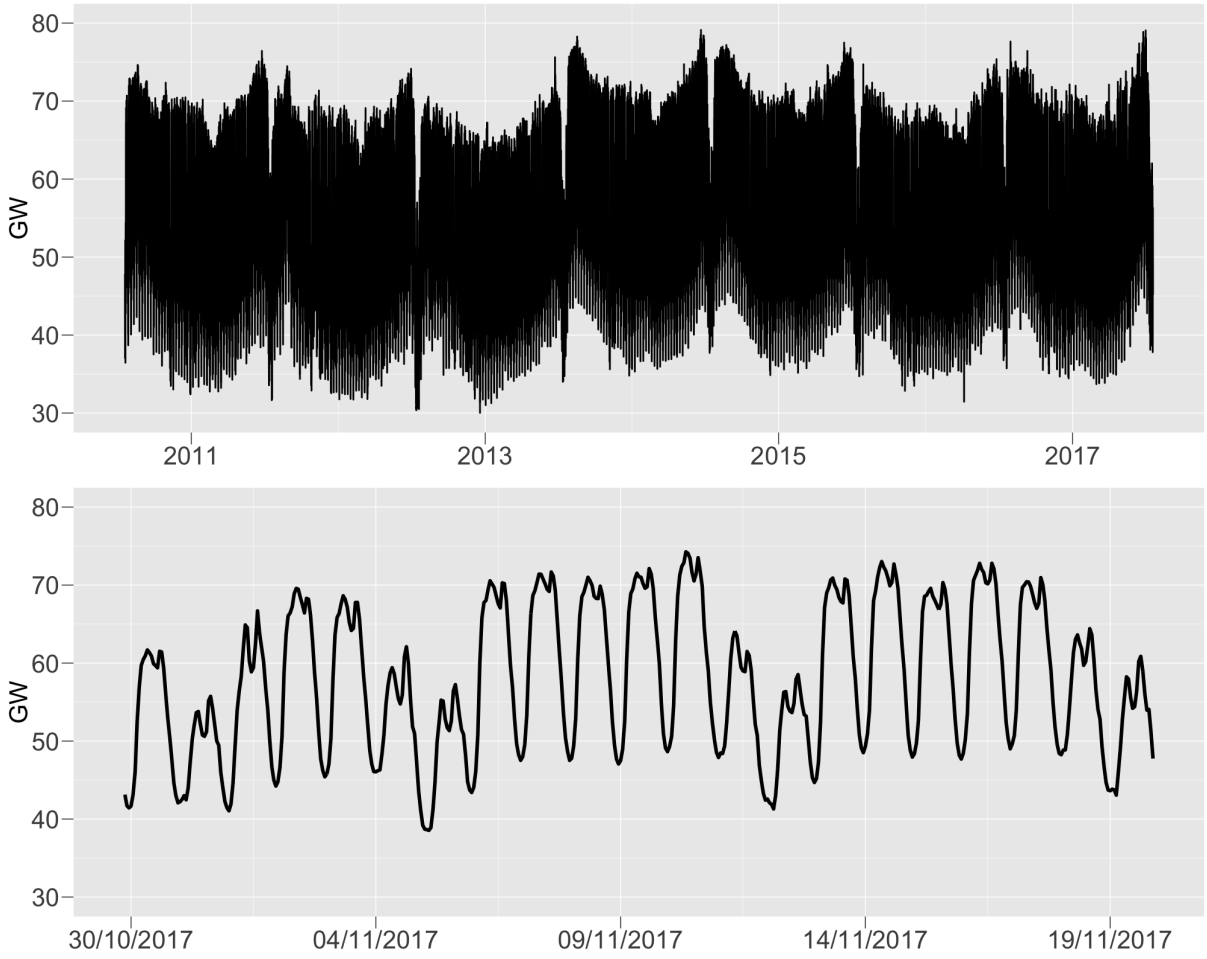

Figure 1.1: Total load in Germany from 2011 trough 2017 (upper) and three consecutive weeks in 2017 (lower) (ENTSO-E (2019)).

to significantly increase the share of renewable energies in the electricity supply. In 2008, this target was first specified and a minimum of 30% of renewable energies on the total gross electricity generation was set for 2020. This was already achieved in 2017. The present targets under the 2014 amendment are shown in Table 1.2.

Table 1.2: Targets for the total gross electricity generation by renewable energy sources according to EEG (2017).

2025	2035	2050
40% - 45%	55% - 60%	min. 80%

In order to support the expansion of renewable energies, three regulations have been anchored in the EEG from the beginning:

- Grid system operators must **connect** installations to generate electricity from re-

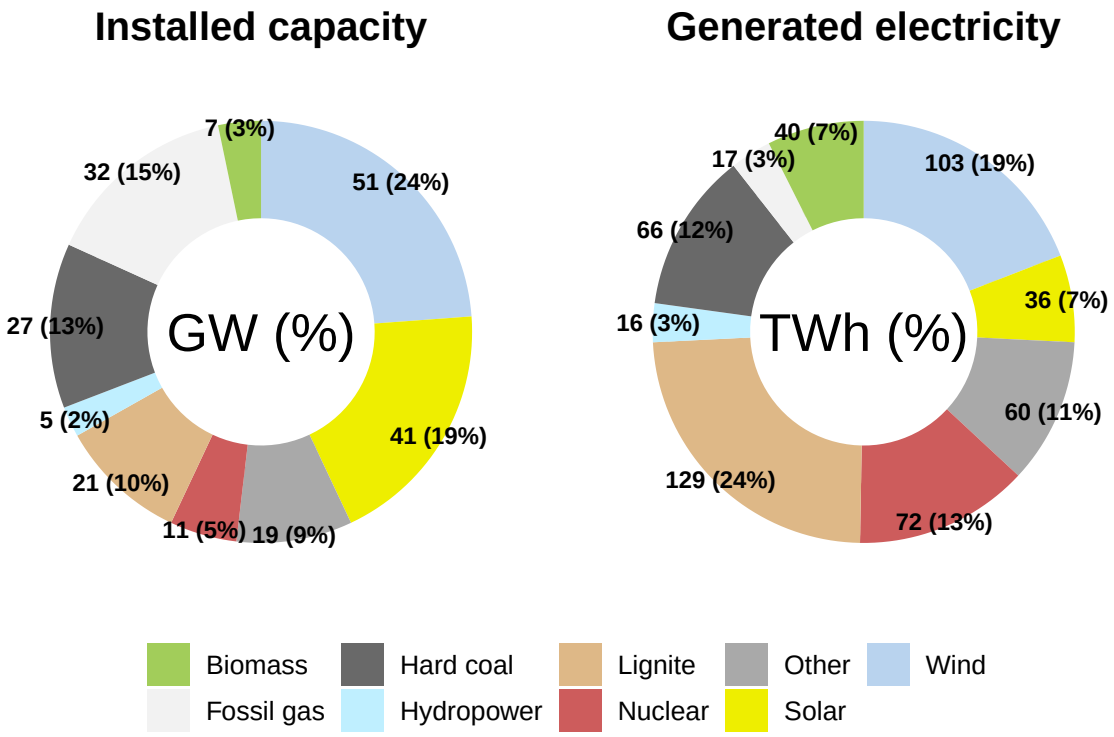

Figure 1.2: Installed capacity (left) and generated electricity (right) in Germany for 2017 by technology (BNetzA (2019)).

newable energy sources **without delay and as a priority**.

- Grid system operators must **purchase**, **transmit** and **distribute** all electricity from renewable energy sources **without delay and as a priority**.
- Operators of installations in which only renewable energy sources are used are **entitled to be paid** for the electricity fed into the grid according to certain **remuneration rates**.

The type of remuneration claims for electricity from renewable sources has evolved over the years from a fixed payment to a market-oriented subsidy. While a fixed payment provides planning security for investors, it does not set an incentive for demand-based production. To ensure that the future expansion of renewable energy takes place economically efficiently, in 2012 it was made possible to bring the generated electricity to the exchange on its own and not to sell directly to the transmission system operators. Since 2016, for newly integrated renewable power plants, this has been mandatory. In order to reduce the resulting price risks, operators receive now a market premium,

$$\text{market premium} = \text{subsidized value - average market price.}$$

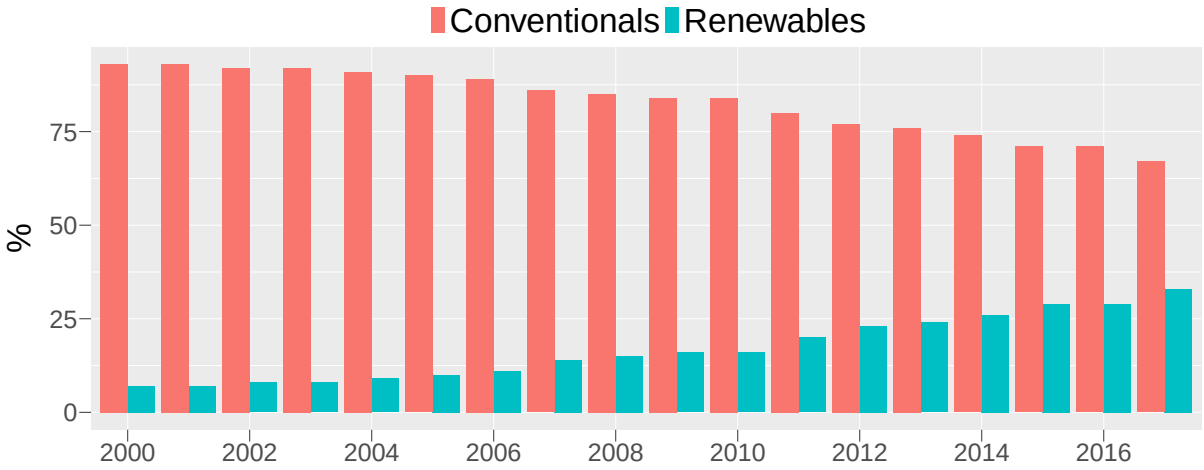

Figure 1.3: Share of renewable and conventional production on total gross electricity generation in Germany (AGEB (2018)).

This support scheme encourages targeted production, since additional revenues can be achieved by adjusting the sales volume to the market price. The objective is to have a personal average sales price above the average market price. The subsidized value has been determined by auction since the EEG 2017. In a pay-as-bid process, investors have the opportunity to secure a share of the tendered promotion volume for the next 20 years.

The support under the EEG has led to a strong increase in renewable power generation in Germany, where solar and wind power have become the key energy sources. Due to their strong expansion (cf. Figure 1.4), they now account for a significant share of the total installed capacity and the generated electricity (cf. Figure 1.2). However, this

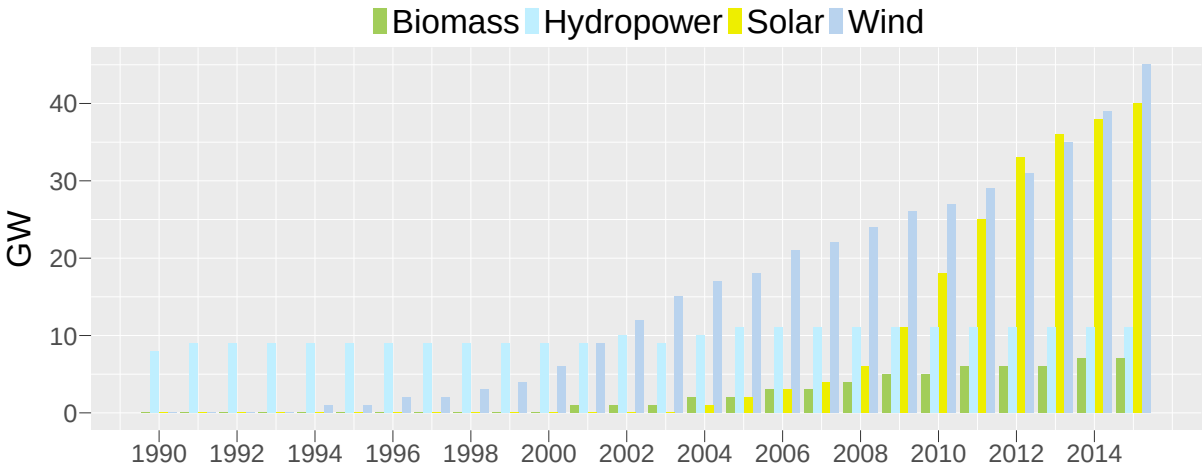

Figure 1.4: Development of installed renewable energy capacity from 1990 to 2015 (Eurostat (2019)).

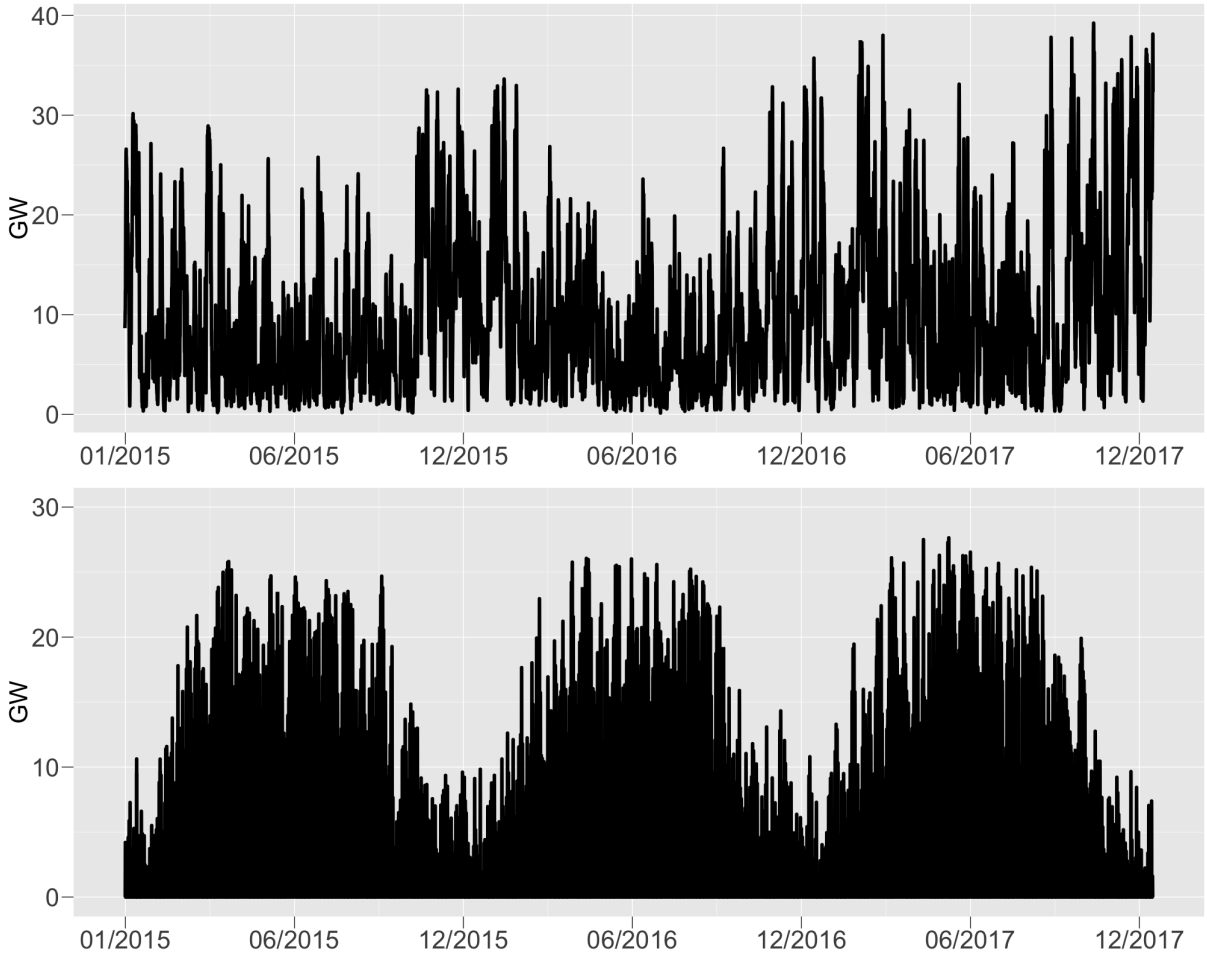

Figure 1.5: Time series of total wind (upper) and total solar power generation (lower) in Germany from 2015 to 2017.

also presents market participants with new challenges. Due to their variable generation, especially solar and wind power cause fluctuations in the supply of electricity. Figure 1.5 shows the large variations of electricity infeed from these two energy sources over the year. Strong variations are not only seasonal, but can also occur within short periods of time (cf. Figure 1.6). Obviously, especially for solar power there is a strong intraday infeed pattern. The variable supply is of great importance for electricity markets. Since their committed exchange trading in Germany, they have at least a strong impact on the price.

1.3.4 The wholesale market

The liberalization of the German electricity market has created the need for a reliable trading platform. Therefore, in August 2000, the EEX was launched. In addition to

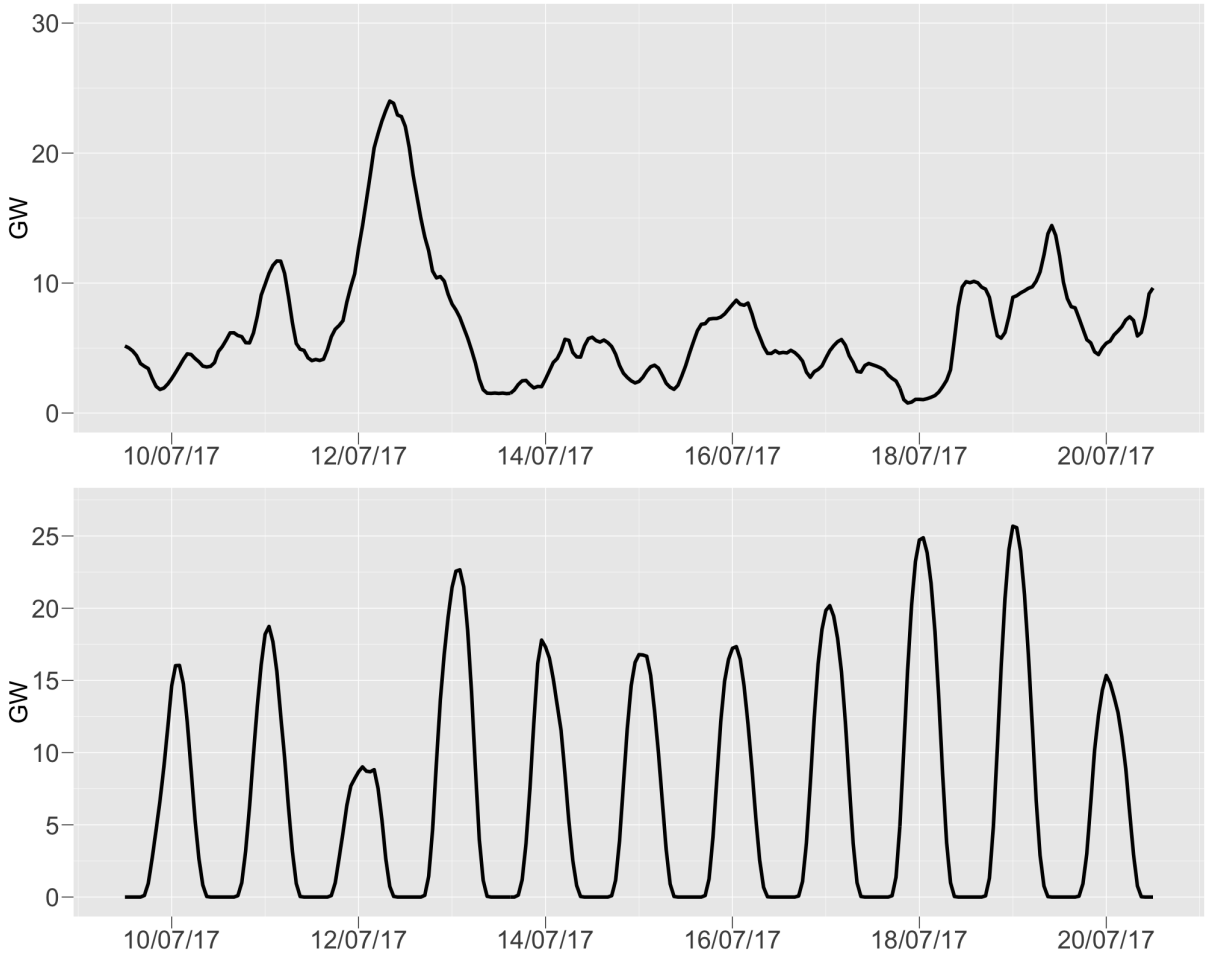

Figure 1.6: Time series of total wind (upper) and total solar power generation (lower) in Germany for eleven consecutive days.

bilateral wholesale trading (OTC trading), EEX offers a reliable marketplace for trading electricity. The most important segments are the spot market, the intraday market and the derivatives market. Short-term transactions are settled via the European Power Exchange (EPEX SPOT), in which EEX holds a major stake.

Spot market The dominant market for short-term electricity trading is the spot market. The spot market is structured in the form of a power exchange (cf. Weron (2006)[Section 1.2]). In 2017, almost half of the total electricity demand in Germany was traded via the spot market (cf. Figure 1.7). In a two-sided auction at 12 noon, bids containing the asked volume and a price per unit can be placed for delivery at a specific hour on the following day. So the spot market is actually a day-ahead market. Delivery must be made physical. The bids are used to construct supply and demand curves for each hour. The intersection of these curves determines the hourly market clearing price (MCP), the price which balances supply and demand. The MCP is set in an uniform-price auction,

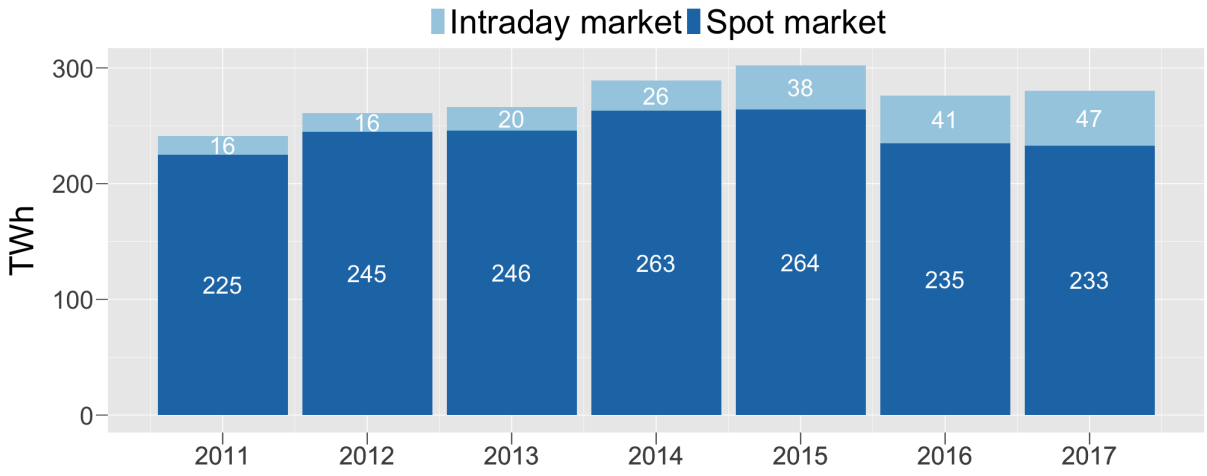

Figure 1.7: Development of spot and intraday market volumes at the EPEX SPOT (BNetzA (2018)).

i.e. suppliers with bids not exceeding the clearing price receive the clearing price, while buyers with bids at least as high as the clearing price pay only this price. Figure 1.8 shows the results of the daily auctions for the years 2016 and 2017.

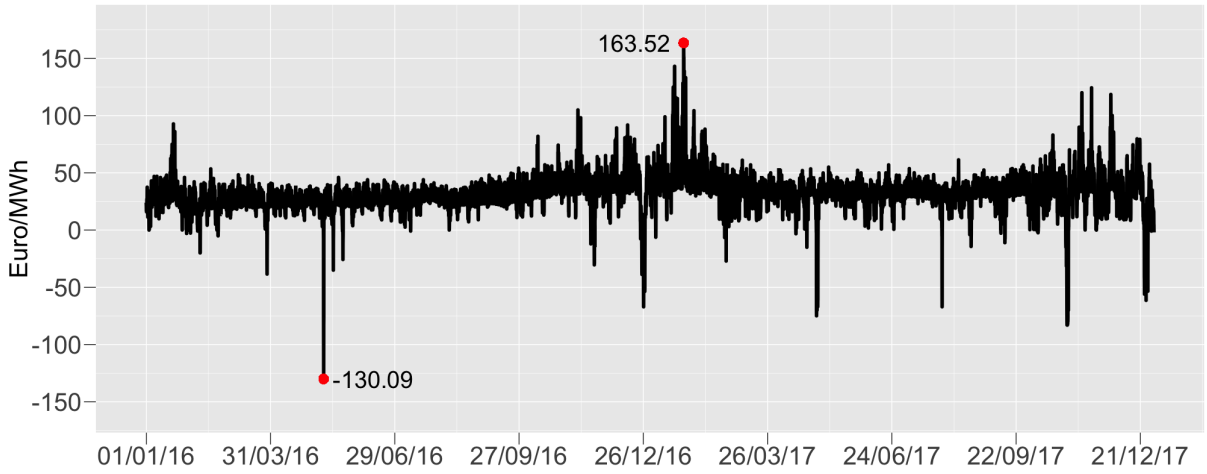

Figure 1.8: EPEX SPOT hourly day-ahead auction results for the years 2016/2017.

Intraday market Very short-term changes to the electricity portfolio can be carried out on the intraday market. To make this possible, the price is not determined by a single auction, in contrast to the spot market. Trading is possible up to 30 minutes before delivery and starts at 3:00 p.m. for hourly contracts of the following day. From 4:00 p.m. also 15-minute periods are tradable. Delivery is physical. In Figure 1.9 the ID_3-Price of the hourly contracts on the German intraday market is illustrated. The ID_3-Price is an index, which represents the volume-weighted price average of all trans-

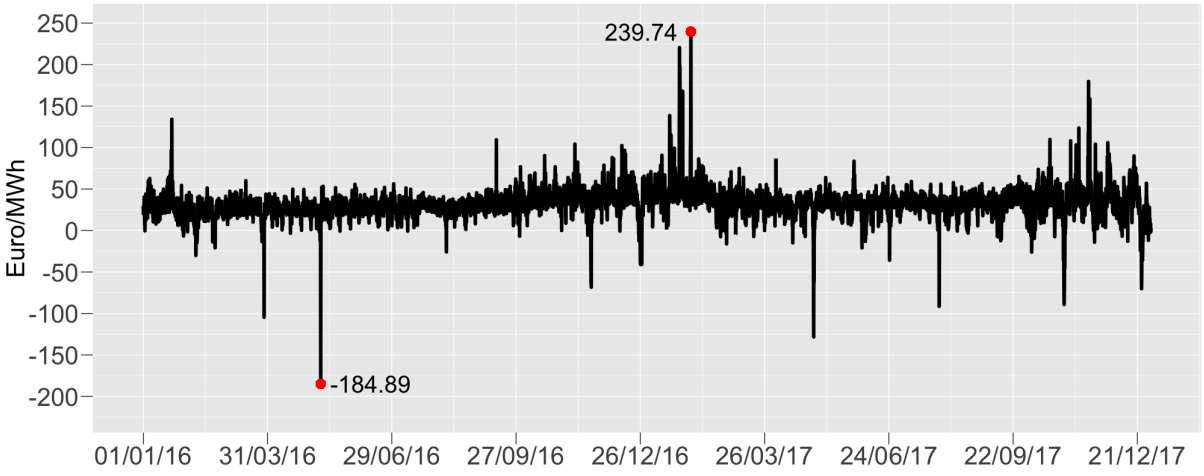

Figure 1.9: ID_3-Price of the hourly contracts on the intraday market for the years 2016/2017.

actions of a traded product performed over the last three hours before start of delivery. The index thus serves as a price indicator for trades that take place shortly before delivery. Such corrective transactions are particularly necessary in an environment with high volatile infeed from renewable energies. The expansion of renewable energy sources in Germany has therefore led to a continuous increase in the importance of the intraday market (cf. Figure 1.7). In 2017, the volume traded again increased considerably by 15% compared to 2016. Due to the further growth of renewable energies, it can be assumed that the intraday market will become even more important.

Derivatives market For mid- and long-term hedging, products on the derivatives market are used. The most important product is the Phelix Future. In 2016, a total volume of 1444 TWh was traded (BNetzA (2018)). There are various delivery periods available of up to six years in the future. However, contracts with delivery more than two years in advance are of little importance (only 12 % of the total trading volume in 2017 (BNetzA (2018))). These derivatives are financially settled contracts with the Phelix as underlying. The Phelix (cf. Figure 1.10) represents the daily arithmetic average of the auction results on the spot market. Options on Phelix Futures can also be traded for individual delivery periods. However, there have been no such transactions for the German market in recent years (BNetzA (2018)).

In response to the challenges for electricity trading arising from the high share of renewable production, in 2015, the EEX also introduced the segment of Energiewende products. This contains new derivatives, specifically designed to cope with the new risks. Currently two products are available. The *German Intraday Cap/Floor Futures* can be used to

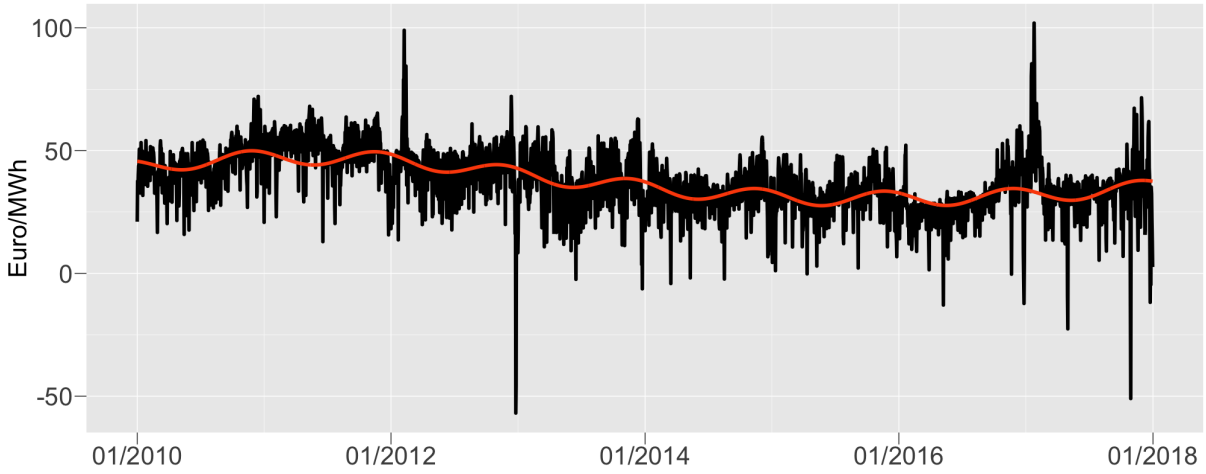

Figure 1.10: Physical electricity index from 2010 to 2017 with approximation of long-term behavior and annual seasonality (red).

hedge against price peaks on the intraday market that frequently occur as a result of high or low generation of renewable energy. While the derivatives presented so far aim to hedge against price risks, the second Energiewende product, the Wind Power Futures, serve to hedge against pure volume risks. For this purpose, the futures have the average load factor of wind power production in Germany as underlying. Since the operator of a wind power plant, subsidized under the EEG, receives a fixed premium, he is no longer exposed to any price risk. This is why futures for renewable energies that only reflect volume risks are reasonable. Due to the strong expansion of renewable energies, it can be assumed that the segment of Energiewende products will continue to gain importance. Therefore, we consider such products in Chapter 3 and Chapter 4.

1.4 Structural price modeling

The replacement of the state electricity price monopoly as a result of liberalization called for market participants to redesign their decision-making mechanism. Electricity price models have become an import input here. We will discuss the characteristics of the pricing process in more detail now and explain their connection to renewable power generation. Weron (2006) is a good source for a comprehensive statistical analysis of electricity prices. We also provide an overview of recent models that define the electricity price as a function of renewable energies. Benth et al. (2008) is the common reference on electricity price modeling. A good survey of different methodologies can be found in Weron (2014).

1.4.1 The merit order

When constructing a price model, it is helpful to understand the underlying market mechanisms. For liberalized electricity markets, prices are often explained using the merit order model. The merit order is the arrangement of suppliers according to their asked prices. These prices should reflect the production costs per generation unit and can approximately be set equal to the marginal costs. At the same time, in accordance with the fixed tariffs of households, a price-inelastic demand is assumed. Since the marginal costs are essentially a matter of the technology used, the merit order can be represented as in Figure 1.11. According to the merit order model, the MCP is the marginal cost of

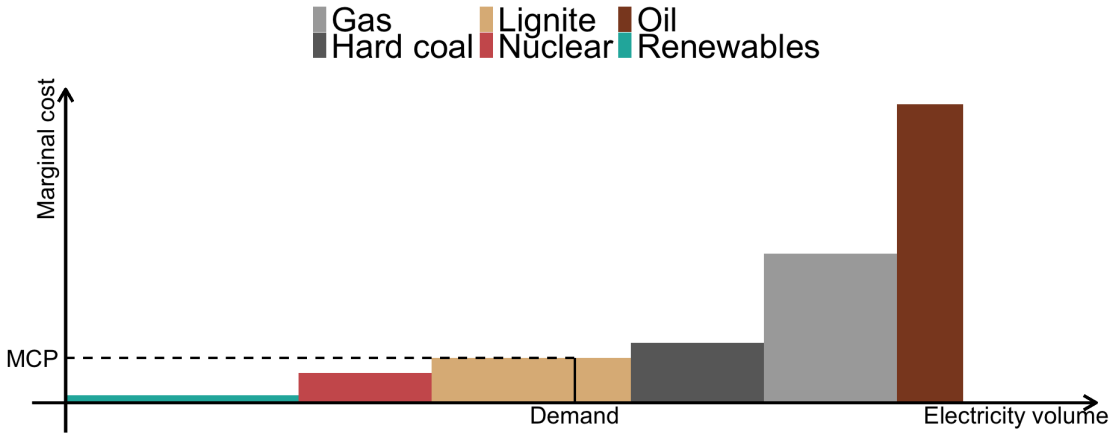

Figure 1.11: Determining the MCP according to the merit order model.

the most expensive power plant needed to meet the demand.

In principle, the price on the spot and intraday market can both be explained using the merit order model. However, due to longer start-up times of some power plants, the available supply decreases as the delivery approaches. Therefore, in the intraday market, a steeper merit order curve can be assumed (cf. Graeber (2014)). This can also be seen by comparing Figure 1.8 and Figure 1.9, where the ID_3-Price shows more extreme values than the day-ahead auction results.

Since electricity from renewable energy sources is produced at vanishing marginal costs, they shift the supply so that cheaper technologies can be used to meet demand. This price-reducing effect of renewable energies can also be seen in Figure 1.10. With the introduction of the market premium model in EEG 2012, average spot market prices fell until 2017.

1.4.2 Stylized facts

The merit order model provides an explanation for the basic level of the electricity price. The real prices, however, are much more complex than this simple model could account for. For electricity price modeling it is therefore also important to consider the real features of the price process. These characteristics are called stylized facts as they are shared by most electricity markets.

Seasonality In Section 1.3 we have already noticed that electricity demand has a strong cyclical behavior. Since the electricity price according to the merit order model is directly connected to demand, prices are cyclic as well. Approximating the annual seasonality by a sinusoid with a period of one year (cf. Figure 1.10), the mirroring of the demand cycle with high values in winter and lower values in summer is demonstrated. Trigonometric functions are used in many applications to describe the seasonality of electricity prices.

Figure 1.12 shows that other periodic patterns of demand are also transferred to the price. Lower prices at night and on weekend can be observed. In addition, the "double

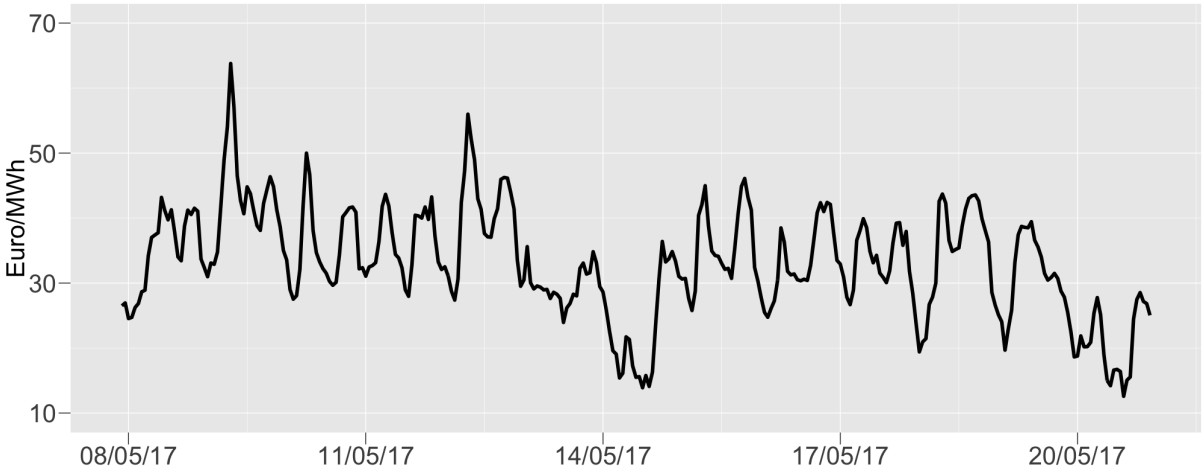

Figure 1.12: Typical cyclic behavior of the electricity price for a period of two weeks using the ID_3-Price as an example.

peak structure" can be detected again. However, this is much more pronounced than for demand. That is because of the price-reducing effect of solar power, which has the highest infeed around noon.

Mean reversion Commodity prices are generally known to be mean-reverting due to the substitution effect. While the return to the mean value can take a long time for commodity prices, a much higher mean-reversion rate can be found in the electricity market. In particular, extreme price levels only persist for a very short time. The price

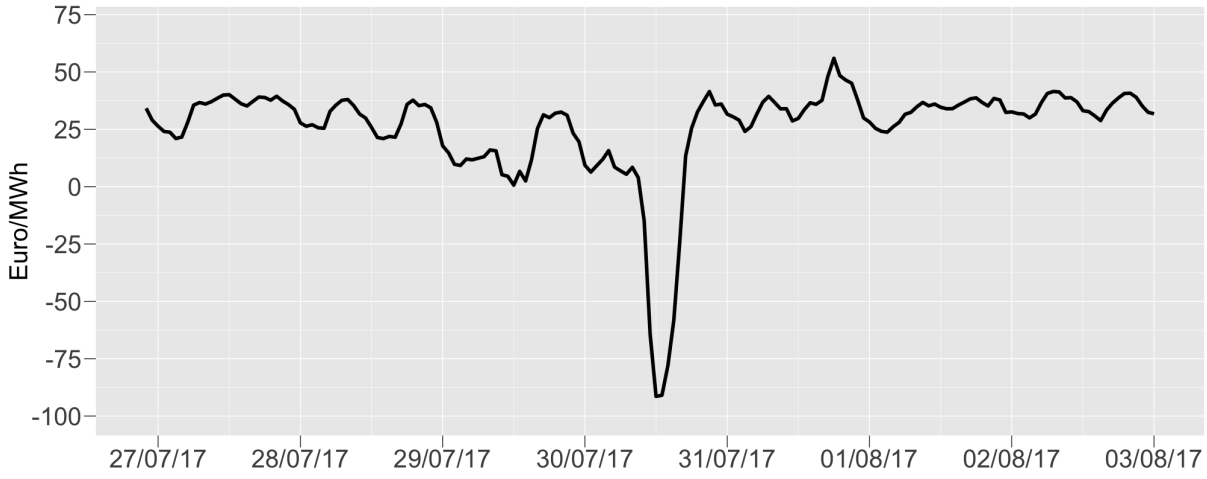

Figure 1.13: Negative ID_3-Price spike on July 30, 2017 at 2:00 pm.

returns to its previous level in a few hours after a jump (cf. Figure 1.13). The fast return also applies to less pronounced deviations from the average level.

The mean-reversion property of electricity prices can again be traced back to the market fundamentals. Many determinants of the market itself are cyclical and mean-reverting. These include, for example, temperature, which has a strong impact on the level of demand, but also the sun and wind intensity, which determine the infeed from renewable energies. For the conventional supply stack, the repair/replacement of a failed power plant can also lead to a normalization of the price.

Jumps and spikes While a strong mean-reversion ensures that a extreme price level is only of short duration, the origin of such a price, that is a jump, is actually the most characteristic feature of electricity markets. Price jumps are abrupt and unexpected extraordinary price changes. Figure 1.13 shows an example, where the ID_3-Price drops almost 95 € within three hours. This is more than the 10-fold of the usual price movement within such a short time. Jumps not only lead to negative but also positive price escalations. Thus, on 31 January 2017, a ID_3-Price of 239 € is temporarily quoted (cf. Figure 1.9), which corresponds to seven times the average price.

Generally, much higher volatility can be observed in electricity markets than for other commodities (cf. Table 1.3). An explanation for strong electricity price movements is given by the high nonlinearity of the merit order curve (cf. Figure 1.11). When demand is very strong, generation at high marginal cost is necessary. Expensive power plants then set the market price. Similarly, an unexpectedly low infeed of renewable energy may require the mobilization of capacities with short start-up times but higher production costs to cover the supply gap. The opposite applies in the case of low demand or high

Table 1.3: Annualized volatility of the relative absolute price changes $|P(t)-P(t-1)|/\bar{P}$, where $\bar{P}$ is the mean, for different markets.

Gas	Oil	Phelix	Day-ahead auction	ID_3-Price
11%	26%	330%	1100%	1353%

renewable energy generation, when only producers with very low costs are in use. Thus the price peak in Figure 1.13 is due to a market situation with a low load on Sunday of about 50 GW paired with a strong generation of solar and wind energy. At peak times, there was only a need for 9 GW of conventional electricity.

However, the cost of production alone cannot explain the extreme prices actually observed. The true reason for these prices is the bidding strategy of the market participants. Since electricity is an essential commodity for some buyers, they are willing to pay virtually any price. Especially since they know that extreme price levels are short-lived. At the same time, producers who cannot flexibly control electricity generation are more willing to accept low or even negative prices at short term rather than face high operational cost. Since both sides are aware of these needs, they adapt their bids accordingly which leads to the observed prices.

1.4.3 Electricity price models

Various attempts have been made to take the stylized facts into account in electricity price modeling. As many of the features can be attributed to the merit order model, early attempts were made to explain the electricity price based on supply and demand.

One of the first concepts was given by Barlow (2002). For a supply $u(t,x)$ and a demand $d(t,x)$ at time t and price x, the electricity price $S(t)$ was defined as the solution of

$$u(t,S(t)) = d(t,S(t)).$$

To solve this equation, quite simple forms for u and d were used. First, u was set equal to the Box-Cox transformation. Then, the inverse of this transformation was applied to an Ornstein-Uhlenbeck (OU) process independent of the price to reflect the price inelasticity of demand.

Despite the simple procedure, the model exhibits price spikes, one important characteristic of electricity prices. However, the definition of the price according to Barlow (2002) is only economically motivated, since the adjustment takes place to pure price data. Nevertheless, in the same spirit as Barlow (2002), other models have been designed which define the price of electricity as the result of fundamental drivers. Due to the

increasing infeed of renewable energies, in recent years there has been special interest in models that can handle information related to renewable power generation (e.g. solar radiation, wind speed, temperature).

Residual demand model Wagner (2012) was one of the first, who defined an electricity price model that includes information on renewable energies. He presented the electricity spot price $S(t)$ at time $t \in [0, T]$ as a deterministic function f of residual demand $R(t)$, i.e.

$$S(t) = f(R(t)),$$

where the residual demand was defined as

$$R(t) = \text{(stochastic) total load} \; - \; \text{(stochastic) infeed from renewables.}$$

Thus the price shift caused by renewable energies was accounted for as a reduction of demand. The function f could therefore be interpreted as the conventional supply stack. For f a hyperbolic function with bounds equal to the price boundaries at the spot market was chosen. This function was fitted based on observed prices and residual demand.

Besides mapping residual demand to the spot price by the function f, Wagner (2012) provided individual models for load, solar and wind power generation. These were used to characterize the residual demand and fitted by appropriate data. Within the renewable energies, only electricity generation from solar and wind power plants was taken into account as they have the largest installed capacities and uncertain infeed profiles.

To specify the behavior of each stochastic component a deterministic seasonality function and an OU process (cf. Section 1.5) were utilized. In other words, the behavior of each fundamental driver abbreviated by driv was determined by

$$\tilde{I}^{\text{driv}}(t) = \eta^{\text{driv}}(t) + \bar{I}^{\text{driv}}(t), \tag{1.1}$$

where

- $\eta^{\text{driv}}(t)$ is a deterministic seasonality function, and
- $\bar{I}^{\text{driv}}(t)$ is an OU process.

For the total load the function $\eta^{\text{load}}(t)$ was specified as a purely periodic function calibrated on the daily average peak demand on business days. Afterwards a deterministic transformation was chosen to take into account intraday variation, holidays and weekend effects.

Due to the strong growth of installed capacities (cf. Figure 1.4), for solar and wind power generation, besides seasonality, also a trend can be observed. To account for this, infeed efficiency $E^{\text{driv}}(t)$ at time t was modeled instead of the absolute values,

$$E^{\text{driv}}(t) = \frac{AI^{\text{driv}}(t)}{IC^{\text{driv}}(t)}, \tag{1.2}$$

where

- $IC^{\text{driv}}(t) > 0$ denotes installed capacity, and
- $AI^{\text{driv}}(t) \in [0, IC^{\text{driv}}(t)]$ is the absolute infeed.

Since $E^{\text{driv}}(t) \in [0,1]$, the logit function, $\text{logit} : (0,1) \to \mathbb{R}$, $x \mapsto \log(\frac{x}{1-x})$, was applied to transfer the wind data to the whole real axis,

$$\tilde{E}^{\text{driv}}(t) = \text{logit}(E^{\text{driv}}(t)).$$

Then the logit wind power efficiency $\tilde{E}^{\text{wind}}(t)$ was modeled according to (1.1).

Since solar infeed is zero at night, transforming the time series to a stationary one is more difficult than for wind. To overcome this problem, observing that the daily infeed patterns look quite similar (cf. Figure 1.5), the daily maximum process $\tilde{M}_i^{\text{solar}}$ of the complete data was considered instead,

$$\tilde{M}_i^{\text{solar}} = \text{logit}\left(\max_{t:d(t)=i} E^{\text{solar}}(t)\right), \ i = 1, 2, \ldots, d(T),$$

where $(d(t) = i) \equiv (t \text{ is on the i-th observed day})$. For $\tilde{M}_i^{\text{solar}}$ again (1.1) was applied. Subsequently, to finally get an intraday solar infeed, functions δ_m, $m = 1, 2, \ldots, 12$, $\delta_m : \{t : m(t) = m\} \times (0,1) \to [0,1]$ have been employed, where $m(t)$ accounts for the month at time t of an observed year in a similar way as $d(t)$ accounts for the days. These functions, called *daily pattern transformations*, were defined as

$$\delta_m(t, x) = x \sum_{k=1}^{24} c_{m,k} \mathbf{1}_{\{t \in (\text{k-th hour of the day})\}}, \tag{1.3}$$

with $c_{m,k}$, $k = 1, 2, \ldots, 24$, set for each month m as the average of

$$\frac{\text{absolute infeed at hour k}}{\text{daily maximum infeed}}.$$

The model of Wagner (2012) makes it easy to consider the impact of renewable energies

on electricity prices. The characteristic features of the spot price are also properly produced. However, for the individual stochastic components it was intended to describe only the most important characteristics. Especially the stochastic behavior may require a more complex model than an OU process. So for the wind power process it was already noted that there is significant partial autocorrelation beside the first lag. Also, as already pointed out by Wagner (2012), all stochastic components are stationary, leaving out long-term effects such as changed fuel prices. Hence additional factors should be added. We will address some of these points in Chapter 3 and Chapter 4.

LSS process Using a regime-switching Lévy semistationary (LSS) process, Veraart (2016) incorporates information on wind power infeed directly into the price process. This is a substantially different concept to the approach of Wagner (2012) where information about renewable energies is mapped to the price through regression.

For Phelix quotations $S(t)$ at time $t \geqslant 0$, Veraart (2016) assumed an arithmetic model,

$$S(t) = \Lambda(t) + Y(t),$$

where $\Lambda(t)$ is a deterministic trend and seasonality function and $Y(t)$ is a LSS process of the form

$$Y(t) = \int_{-\infty}^{t} g(t-s)dM(s),\ t \geqslant 0.$$

Here, $g : \mathbb{R} \to \mathbb{R}$ is set to the kernel of a CARMA(2,1) process (cf. Section 1.5) and

$$dM(s) = \rho(s)dM^{(1)}(s) + (1-\rho(s))dM^{(2)}(s),$$

for independent Lévy processes $M^{(1)}(t)$ and $M^{(2)}(t)$. This model allows for two different behaviors of the price, depending on whether a synthetic wind penetration index is high or low by altering $\rho(s)$ according to

$$\rho(s) = \begin{cases} 1, & \text{if the wind penetration index at time s is “high”,} \\ 0, & \text{if the wind penetration index at time s is “low”.} \end{cases}$$

The wind penetration index $WP(t)$ was defined as the ratio between the daily forecast of wind power generation $V(t)$ and the daily forecast of load $L(t)$,

$$WP(t) = \frac{V(t)}{L(t)} \in [0,1], \tag{1.4}$$

since it was found out, that rather extreme negative day-ahead prices are associated with

a high value of this index. Using forecast information, this index could be calculated. For fitting the LSS process, first the CARMA parameters have been identified. Afterwards, the increments of the driving process have been filtered and separated according to the first quartile of the wind penetration index. An individual generalized hyperbolic (GH) distribution was then fitted to each set of increments.

The use of an LSS process allows to incorporate information on renewable energies in a continuous-time stochastic model in a flexible way that goes beyond pure regression. Obviously, it would be desirable to consider other energy sources, in particular solar power, in addition to wind power. Furthermore, Veraart (2016) does not develop a stochastic model for the fundamental drivers, which would be necessary to analyze the mid- and long-term price level. In addition, the choice of the switching variable seems relatively arbitrary, especially since it was shown that the behavior of the increments cannot be strictly divided into two classes. We deal with these points in Chapter 4.

Cox processes Doubly stochastic Poisson or Cox processes represent a well-known class of stochastic processes to handle external information. Deschatre and Veraart (2017) used a Cox process with intensity as a function of the wind penetration index $WP(t)$ (cf. Equation (1.4)) to model the occurrence of spot price jumps. They found that the probability of an extreme negative return increases with the value of the index.

Hence, they decomposed the spot price $S(t)$ at time $t \in [0, T]$ in the following way:

$$S(t) = \Gamma_1(t) + X_1(t) + Y^+(t) + Y^-(t),$$

where $\Gamma_1(t)$ is a deterministic trend and seasonality function, $X_1(t)$ is a CAR(24) process (cf. Section 1.5) and $Y^{\pm}(t)$ satisfy

$$Y^{\pm}(t) = \int_0^t \int_{\mathbb{R}} x e^{-\beta(t-s)} p^{\pm}(ds, dx).$$

Here, $\beta > 0$ is a common speed of mean reversion and $p^{\pm}(dt, dx)$ are Poisson measures with compensators $q^{\pm}(dt, dx) = \lambda^{\pm}(t)dt \times \nu^{\pm}(dx)$ for finite measures $\nu^{\pm}$ satisfying $\nu^{\pm}(\{0\}) = 0$ and $\int_{\mathbb{R}} x^2 \nu^{\pm}(dx) < \infty$. For positive jumps $\lambda^+(t)$ was set to a constant. To make the intensity $\lambda^-(t)$ of the negative jumps dependent on wind penetration, they assumed the functional form

$$\lambda^-(t) = q(WP(t)),$$

where $q : [0, 1] \to (0, \infty)$ is a deterministic function found by nonparametric regression. The distribution of the jump sizes was not specified. Deschatre and Veraart (2017) also

established a stochastic model for the wind penetration index,

$$\text{logit}(WP(t)) = \Gamma_2(t) + X_2(t),$$

where again $\Gamma_2(t)$ is a deterministic trend and seasonality function and $X_2(t)$ is a CAR(24) process.

The model of Deschatre and Veraart (2017) maps information on wind power and load to the spot price in a complete stochastic framework. Of course, also other renewable energies should be considered. In contrast to the LSS process of Veraart (2016), however, for the Cox process the distribution is allowed to depend on the wind penetration in a continuous way. But this only applies to the jump times of the spot price. In general, the jump sizes cannot be controlled by applying a Cox process. That this is also advisable will be shown in Chapter 4.

Stochastic time change Schmeck and Borovkova (2017) modeled Phelix quotations $S(t)$ at time $t \geqslant 0$ with respect to temperatures $C(t)$ in Cologne. Temperature as a fundamental driver of the price seems reasonable as there is strong correlation to demand, e.g. through electric heating.

Since the variations of the electricity price are not similar at each calendar day, the idea of Schmeck and Borovkova (2017) was to choose a time grid for which the behavior is more alike. For this purpose, a time change $T(t)$ was defined by

$$T(t) = \int_0^t \tau(s)\, ds,\ \tau(t) = \log(1 + g(C(t))) + k,$$

with g a polynomial that maps temperature to demand and k as a normalization constant.

Then, Schmeck and Borovkova (2017) postulated that the price is given by

$$S(t) = s(t) + Y(t),$$

where $s(t)$ is a deterministic seasonality function and for $Y(t)$ holds

$$Y(t) = \mu\left(1 - e^{-\theta T(t)}\right) + \int_0^t e^{-\theta(T(t)-T(s))}\sigma\sqrt{\tau(s)}dW(s) + \int_0^t e^{-\theta(T(t)-T(s))}dZ(s),$$

with $W(t)$ a Brownian motion and $dZ(t) = \int_{\mathbb{R}} xN(dx, dt)$ for an integer-valued random measure $N(dx, dt)$ with compensator $\nu(dx, dt) = \tau(t)\lambda dt \times f(x)dx$. The process $Y(t)$ is the result of time-changing an OU process $X(t)$ with jumps, i.e. $Y(t) = X(T(t))$. Here,

θ, μ and σ are the OU parameters (cf. Section 1.5) and $X(t)$ is driven by the Brownian motion $W(t)$ and a compound Poisson process with jump intensity $\lambda > 0$ and f the density of the jump size distribution.

For the temperature $C(t)$ an arithmetic model consisting of the sum of a deterministic seasonal component and an OU process was proposed. The OU parameters of $Y(t)$ were fitted by ordinary least squares using the discrete-time representation of the process. The jumps were filtered prior to that and handled separately.

The time change suggested by Schmeck and Borovkova (2017) leads to a process of OU type with a speed of mean-reversion, variance and jump intensity depending on temperature. This enables the modeling of seasonal patterns in electricity price volatility and spikes. In Germany in winter the general price volatility is higher and price jumps are more frequent than in summer (cf. Figure 1.8). One drawback using a time-change is that the jump size distribution cannot be altered. Obviously other fundamental drivers, especially renewable energies, should be considered to control the time change. Besides these points, the greatest challenge of this approach lies in the choice of the time change itself. Since price data is not directly used for its selection, it can be assumed that the relationship between price and fundamental drivers is not fully described. Hence, in Chapter 4 we will use a more direct method to map this dependency.

1.5 Continuous-time autoregressions

Originally defined for applications in physics, there is now an interest in continuous-time autoregressive (CAR) processes in many other areas as well. Since we frequently use these type of processes in this thesis, we summarize here some of their essential properties. A more detailed review of continuous-time autoregressive processes is given by Brockwell (2014). The following sections assume a knowledge of stochastic processes and stochastic calculus as it can be found in any standard work. For Subsection 1.5.1, e.g. Karatzas and Shreve (1998) provides sufficient information, while Protter (2005) and Jacod and Shiryaev (2002) are common references for more general definitions of stochastic integration.

1.5.1 The OU process

The best known member of the class of CAR processes is the OU process. The OU process is defined by the stochastic differential equation (SDE)

$$dX^{\mu}(t) = \theta(\mu - X^{\mu}(t))dt + \sigma dW(t),$$

where $\theta,\ \mu \in \mathbb{R},\ \sigma > 0$ are constants and $W = \{W(t)\}_{t\geqslant 0}$ is a standard Brownian motion. Given an initial value $X^{\mu}(0)$ (a random variable independent of W), the solution $X^{\mu} = \{X^{\mu}(t)\}_{t\geqslant 0}$ to this SDE is

$$X^{\mu}(t) = \mu + (X^{\mu}(0) - \mu)\, e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-u)} dW(u), \quad t \geqslant 0. \tag{1.5}$$

Sometimes the parameter μ is eliminated by considering $X = X^{\mu} - \mu$ rather than X^{μ}.

The OU process is a Gaussian Markov process having continuous paths, which was proposed by Uhlenbeck and Ornstein (1930) to describe the velocity of a particle suspended in a liquid. In finance, the OU process is also known as part of the Vasicek (1977) model, where it was applied for interest rates modeling.

The main attraction of the OU process in statistical modeling stems from its mean-reverting property. For $\theta > 0$, the instantaneous drift $\theta(\mu - X^{\mu}(t))$ guarantees that the process is pulled towards μ. The greater θ and the deviation, the stronger the attraction. Erratic fluctuations around μ with a magnitude set by σ are caused by the Brownian motion. Therefore, θ is also referred to as *mean-reversion speed*, μ as *mean-reversion level* and σ as *instantaneous volatility*.

Mean-reversion also induces a stationary distribution of the OU process. From (1.5) we get

$$X(t+h) = e^{-\theta h} X(t) + \sigma \int_t^{t+h} e^{-\theta(t+h-u)} dW(u),\ t, h \geqslant 0, \tag{1.6}$$

and it is immediately clear (by the independence of the increments of W) that X is a Markov process. Using the Itô isometry the stochastic integral is a Gaussian process with mean 0 and variance $\frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta h}\right)$ for all $h \geqslant 0$. Hence, we can immediately express the transition distribution of the OU process as

$$\mathbb{P}(X(t+h) \leqslant y | X(t) = x) = \Phi\left(\frac{\sqrt{2\theta}(y - xe^{-\theta h})}{\sigma\sqrt{1 - e^{-2\theta h}}}\right), \tag{1.7}$$

where Φ is the standard normal distribution. If $\theta > 0$, the limiting distribution is

$$\lim_{h\to\infty} \mathbb{P}(X(t+h) \leqslant y | X(t) = x) = \Phi\left(\frac{\sqrt{2\theta}y}{\sigma}\right).$$

Hence, if $X(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\right)$, then the limiting distribution is the stationary distribution and holds for all times $t \geqslant 0$,

$$X(t) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\right) \Leftrightarrow X^{\mu}(t) \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{2\theta}\right).$$

The stationary distribution clearly expresses that X^{μ} fluctuates around the level μ. This shows again the mean-reverting property of an OU process.

The stationarity of the OU process should not to be confused with the existence of a stationary distribution. By using (1.6), it is easy to derive that for $\mathbb{E}\left(X^2(0)\right) < \infty$,

$$\mathbb{E}(X(t)) = e^{-\theta t}\mathbb{E}(X(0)),\ t \geqslant 0,$$

and

$$Cov(X(t), X(s)) = \frac{\sigma^2}{2\theta}\left(e^{-\theta(t-s)} - e^{-\theta(t+s)}\right) + e^{-\theta(t+s)}Var(X(0)),\ 0 \leqslant s \leqslant t.$$

If $X(0)$ has mean 0 and variance $\frac{\sigma^2}{2\theta}$, X has constant mean and a covariance function that only depends on the time between the observations. Then, the autocovariance function (ACVF) is

$$\gamma_X(h) = Cov(X(t+h), X(t)) = \frac{\sigma^2}{2\theta}e^{-\theta h},\ h \geqslant 0. \tag{1.8}$$

Therefore, weak-sense stationarity is achieved. If even $X(0) \sim \mathcal{N}(0, \frac{\sigma^2}{2\theta})$ applies, X is also strictly stationary, since the stationary distribution is the Gaussian distribution.

Even if $X(0)$ does not satisfy the above conditions, by (1.6) the influence of the initial value exponentially decreases. Hence, stationarity is at least achieved asymptotically for $\theta > 0$. Also for $\theta \leqslant 0$ the process given by (1.5) is well-defined. Of course, for $\theta < 0$ the stochastic integral "explodes" as $t \to \infty$.

1.5.2 CARMA processes

In physics, the application of the OU process is motivated by Langevin's equation of motion. In other fields it is chosen due to its appealing properties, especially the mean-

reversion. However, two major shortcomings were identified. First, the Brownian motion seems not always appropriate to explain the variability of the data. Second, by (1.8), the stationary OU process has an exponentially decreasing ACVF. But such a simple dependence structure is not always sufficient. Therefore, the more general continuous-time autoregressive moving average (CARMA) process was defined.

A (Lévy-driven) CARMA(p, q) process $Y = \{Y(t)\}_{t\geqslant 0}$ of order $p, q \in \mathbb{N}$ with $q < p$ is defined (cf. Brockwell (2014)) by the equation

$$Y(t) = \mathbf{b}^T \mathbf{X}(t), \tag{1.9}$$

where the $\mathbb{R}^p$-valued process $\mathbf{X} = \{\mathbf{X}(t)\}_{t\geqslant 0}$ is the solution to the SDE

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + \mathbf{1}_p dL(t), \tag{1.10}$$

driven by a Lévy process $L = \{L(t)\}_{t\geqslant 0}$ with

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \dots & -a_1 \end{pmatrix}, \mathbf{1}_p = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \text{ and } \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{pmatrix},$$

for real-valued coefficients $a_1, \dots, a_p$, $b_0, \dots, b_{p-1}$ satisfying $b_q = 1$ and $b_j = 0$, $j > q$. For $p = 1$ the matrix A is to be understood as $A = (-a_1)$. Given an initial random vector $\mathbf{X}(0) \in \mathbb{R}^p$ independent of L,

$$\mathbf{X}(t+h) = e^{Ah}\mathbf{X}(t) + \int_t^{t+h} e^{A(t+h-u)}\mathbf{1}_p dL(u), \ t, h \geqslant 0. \tag{1.11}$$

A CARMA(p,q) processes of order $q = 0$ is also called a CAR process and denoted as CAR(p).

Remark 1.1

The condition $b_q = 1$ ensures the identifiability of the CARMA process and in particular of the Levy process L. Instead of the q-th component, any other b_j, $j \leqslant q$, can also be fixed. In numerical applications $b_0 = 1$ is often preferred. This also connects CARMA processes more directly with their discrete-time counterparts, the ARMA processes.

Comparing (1.6) and (1.11), there is obviously a strong connection between the class of CARMA processes and the OU process. In fact, for $L \equiv \sigma W$ and $p = 1$ we get an OU process with $\theta = a_1$ and $\mu = 0$. Accordingly, many features of the OU process can be assigned to CARMA processes. Of particular interest is again the mean-reversion property.

Mean-reversion For $\mathbb{E}(L(1)) = \xi < \infty$, we get by (1.11),

$$\mathbf{X}(t+h) = -A^{-1}\xi\mathbf{1}_p + e^{Ah}\left(\mathbf{X}(t) + A^{-1}\xi\mathbf{1}_p\right) + \int_t^{t+h} e^{A(t+h-u)}\mathbf{1}_p d(L(u) - \xi u). \quad (1.12)$$

The diagonalization (over $\mathbb{C}$) of $A = P\Lambda P^{-1}$ gives the matrix exponential $e^{Ah} = Pe^{\Lambda h}P^{-1}$. Now, if the eigenvalues of A have strictly negative real part, i.e.

$$Re(\lambda_i) < 0,\ i = 1, \ldots, p, \quad (1.13)$$

where λ_i are the diagonal elements of Λ, the process $\mathbf{X}$ returns to $-A^{-1}\xi\mathbf{1}_p$. Of course, this means that $Y = \mathbf{b}^T\mathbf{X}$ returns to $-\mathbf{b}^T A^{-1}\xi\mathbf{1}_p$ and for $\xi = 0$, $Y^\mu = Y + \mu$ fluctuates around the level μ. Since for the OU process $\lambda_1 = -\theta$, in this case, condition (1.13) is equivalent to the requirement of a positive mean-reversion parameter.

Stationarity Stationarity of $\mathbf{X}$ resp. Y can also be shown under condition (1.13). Since the stochastic integral in (1.12) has mean 0, for weak-sense stationarity we need $\mathbb{E}(\mathbf{X}(0)) = -A^{-1}\xi\mathbf{1}_p$. Furthermore, if $Var(L(1)) = \sigma^2 < \infty$, using (1.11) and the i.i.d increments property of L,

$$Cov(\mathbf{X}(t), \mathbf{X}(s)) = e^{At}\left(Cov(\mathbf{X}(0)) + Cov\left(\int_0^s e^{-Au}\mathbf{1}_p dL(u)\right)\right)e^{As} =$$

$$e^{A(t-s)}Cov\left(e^{As}\mathbf{X}(0) + \int_0^s e^{A(s-u)}\mathbf{1}_p dL(u)\right) =$$

$$e^{A(t-s)}Cov\left(\int_s^\infty e^{Au}\mathbf{1}_p dL(u) + \int_0^s e^{Au}\mathbf{1}_p dL(u)\right) = e^{A(t-s)}\Sigma, \quad 0 \leqslant s \leqslant t,$$

where $Cov(\mathbf{X}(0))$ is assumed to be equal to the covariance matrix Σ of $\int_0^\infty e^{Au}\mathbf{1}_p dL(u)$. As shown above, this is necessary and sufficient to achieve weak-sense stationarity of $\mathbf{X}$. Like for the OU process, $\mathbf{X}$ is also strictly stationary, if $\mathbf{X}(0)$ is distributed according to the limiting distribution $\int_0^\infty e^{Au}\mathbf{1}_p dL(u)$. The same holds for Y, where we get

$$\mathbb{E}(Y(t)) = -\mathbf{b}^T A^{-1}\xi\mathbf{1}_p,\ t \geqslant 0,$$

and

$$\gamma_Y(h) = \mathbf{b}^T e^{Ah}\Sigma\mathbf{b},\ h \geqslant 0.$$

This ACVF allows for much more representations of dependency structures than possible for the OU process. It does not even have to be monotone.

Example 1.2
The matrix A with $a_1 = 1$, $a_2 = \frac{\pi^2+1}{4}$ has eigenvalues $\lambda_1 = -0.5 + i\frac{\pi}{2}$, $\lambda_2 = -0.5 - i\frac{\pi}{2}$. If $X(0) \sim \int_0^\infty e^{Au}\mathbf{1}_p dL(u)$, then, setting $b_0 = 0$ and $b_1 = 1$, leads to the strictly stationary CARMA(2,1) process Y with ACVF,

$$\gamma_Y(h) = \frac{\sigma^2}{2}e^{-0.5h}\left[\cos\left(\frac{\pi h}{2}\right) - \frac{\pi}{4}\sin\left(\frac{\pi h}{2}\right)\right],$$

which exhibits a damped oscillatory behavior.

However, this generalization of the ACVF is paid for with an important property. While $\mathbf{X}$ is a Markov process (cf. Equation (1.11)), this is no longer the case for the CARMA process Y, which is a linear combination of the components of $\mathbf{X}$. But this is inevitable, because according to Doob (1942) every (except of a trivial white noise sequence) strictly stationary, Gaussian, Markovian process is an OU process, i.e. has an exponentially decreasing ACVF.

The loss of the Markov property not only poses theoretical challenges, but also practical ones. While for the OU process we could use simple transition distributions (cf. Equation (1.7)) for simulation and estimation, these are not directly available for CARMA processes.

Estimation A potential solution for estimating the parameters of a strictly stationary CARMA process is given by the fact, that the observations $Y(t_1), Y(t_2), \ldots, Y(t_N)$ at times $0 \leqslant t_1 < \cdots < t_N$ satisfy the equations

$$\begin{aligned} Y(t_i) &= \mathbf{b}^T\mathbf{X}(t_i), \quad i = 2, \ldots, N, \\ \mathbf{X}(t_i) &= e^{A(t_i - t_{i-1})}\mathbf{X}(t_{i-1}) + \mathbf{Z}_i, \end{aligned} \tag{1.14}$$

where $\mathbf{X}(t_1) \sim \int_0^\infty e^{Au}\mathbf{1}_p dL(u)$ and $\{\mathbf{Z}_i\}_{i=2,\ldots,N}$ is a set of independent random vectors with $\mathbf{Z}_i \sim \int_{t_{i-1}}^{t_i} e^{A(t_i-u)}\mathbf{1}_p dL(u)$. These equations precisely define a linear state-space model and the related theory (cf. Durbin and Koopman (2012)) can be applied. Besides the evaluation of the matrix exponential e^{Ah}, $h \geqslant 0$, this requires some knowledge about the distribution of $\mathbf{Z}_i$, $\forall i$. For $L = \sigma W$, it is enough to know the covariance matrices of

$\mathbf{Z}_i$, which can be found from Σ noting that

$$Cov(\mathbf{Z}_i) = \Sigma - e^{A(t_i - t_{i-1})} \Sigma e^{A^T (t_i - t_{i-1})}.$$

An algorithm for calculating Σ given the matrix A is stated e.g. in Tómasson (2015). In case L is a non Gaussian Lévy process, of course, it is not sufficient to know Σ. However, the CARMA parameters a_j, $j \leqslant p$,b_j, $j \leqslant q$, can still be estimated as in the Gaussian case. Schlemm and Stelzer (2012) proved consistency of this quasi-maximum likelihood estimation.

This leaves the problem of choosing a proper Lévy process. Restating (1.10) allows to recover the increments of L from the process $\mathbf{X}$,

$$L(t) - L(s) = \mathbf{1}_p^T \left(\mathbf{X}(t) - \mathbf{X}(s) - \int_s^t A\mathbf{X}(u)du \right), \ 0 \leqslant s \leqslant t. \tag{1.15}$$

For inference concerning the process L, it is therefore enough to identify $\mathbf{X}$. This is possible from realizations of Y as indicated by Brockwell and Lindner (2015). Using the notation $\mathbf{X} = \mathbf{X}_p = (X_0 \ \cdots \ X_{p-1})^T$, from (1.10) and the definition of the matrix A, we have

$$dX_j(t) = X_{j+1}(t)dt, \ j = 0, \ldots, p-2. \tag{1.16}$$

If $q = 0$, $X_0 = Y$ and the remaining $p-1$ components are obtained by successive differentiation. Otherwise, $X_q = b_q X_q = Y - b_0 X_0 - \ldots - b_{q-1} X_{q-1}$. Hence, $\mathbf{X}_q$ satisfies the SDE

$$d\mathbf{X}_q(t) = B\mathbf{X}_q(t)dt + \mathbf{1}_q Y(t)dt,$$

or equivalently

$$\mathbf{X}_q(t) = e^{B(t-s)}\mathbf{X}_q(s) + \int_s^t e^{B(t-u)}\mathbf{1}_q Y(u)du, \ 0 \leqslant s \leqslant t, \tag{1.17}$$

with

$$B = \begin{pmatrix} 0 & I_{q-1} \\ -\mathbf{b}_q^T \end{pmatrix}, \text{ and } \mathbf{b}_q^T = (b_0 \ \cdots \ b_{q-1})^T,$$

where I_{q-1} is the $(q-1) \times (q-1)$ identity matrix. For $q = 1$, the matrix B boils down to $(-b_0)$. Assuming a plausible value for $\mathbf{X}_q(0)$, (1.17) gives a method for filtering $\mathbf{X}_q$ from Y. In order that the pathwise integral in (1.17) for large intervals does not explode, it is necessary that the eigenvalues μ_i, $i = 1, \ldots, q$, of B have negative real part, i.e.

$$Re(\mu_i) < 0, \ i = 1, \ldots, q.$$

Then, according to (1.16), the complete process $\mathbf{X}$ is obtained by differentiation of X_0 to find the remaining components X_j, $q-1 < j \leqslant p-1$. Applying (1.15), the increments of the Lévy process L can be recovered, where the integrals appearing above are numerically approximated in application.

Example 1.3
For a strictly stationary CARMA(2,1) process Y with $b_0 > 0$ driven by the Lévy process L, applying (1.17)*, the components of the process $\boldsymbol{X}$ are*

$$\begin{aligned} X_0(t) &= e^{-b_0 t} X_0(0) + \int_0^t e^{-b_0(t-u)} Y(u)du, \ t \geqslant 0, \\ X_1(t) &= Y(t) - b_0 X_0(t). \end{aligned}$$

Therefore, by (1.15)*, the increments of L can be recovered as*

$$L(t) - L(s) = Y(t) - Y(s) + (a_1 - b_0)(X_0(t) - X_0(s)) + a_2 \int_s^t X_0(u)du.$$

Simulation By simulation many properties of a stochastic process can be made numerically accessible. This is particularly important if analytic solutions are not available. In principle, the linear state-space model (1.14) can be used to generate realizations of a CARMA process. However, this requires the evaluation of the matrix exponential e^{Ah}, $h \geqslant 0$, and a way to simulate the random vectors $\{Z_i\}_{i=2,\dots,N}$. Since especially the last point can be difficult, a more direct procedure is often used. An approximation $\{\mathbf{X}^{(n)}(\tau_i)\}_{i=0,\dots,n}$, $0 = \tau_0 < \cdots < \tau_n = T$, $\tau_i - \tau_{i-1} = \frac{T}{n}$, to $\{\mathbf{X}(t)\}_{0 \leqslant t \leqslant T}$ is given by the discretization of the SDE (1.10), i.e.

$$\mathbf{X}^{(n)}(\tau_i) = \mathbf{X}^{(n)}(\tau_{i-1}) + A\mathbf{X}^{(n)}(\tau_{i-1})(\tau_i - \tau_{i-1}) + \mathbf{1}_p \left(L(\tau_i) - L(\tau_{i-1})\right).$$

This procedure is known as the Euler scheme and consistency ($n \to \infty$) for a Lévy driven SDE has been proven by Jacod (2004). Using a small value for n, this scheme enables to generate realizations of $\{\mathbf{X}(t)\}_{0 \leqslant t \leqslant T}$ given some initial value $\mathbf{X}(0) = x$, and thus also of the related CARMA process Y provided on can simulate the Lévy process L.

Chapter 2

Continuous-Time Threshold Autoregressions with Jumps

In this chapter we study a nonlinear extension of continuous-time autoregressive (CAR) processes that can exhibit jumps. We start with a short survey of nonlinear autoregressive processes. Then we define a nonlinear CAR process with jumps and prove its existence in Section 2.2 using a generalized version of the Girsanov theorem. In Section 2.3 we show that an Euler scheme can be used for approximating this process. A maximum likelihood estimator is constructed in Section 2.4 using particle filtering methods. We test its quality in a simulation study. Finally we apply the process to the Physical Electricity Index in Section 2.5 and show that the inclusion of a nonlinear autoregressive component provides a significant advantage over comparable models.

2.1 Nonlinear CAR processes

Continuous-time autoregressive moving average (CARMA) processes represent an import class in modeling stationary time series. They are used to map a functional relationship between successive observations. However, for the traditional CARMA process discussed in Section 1.5 this relationship is limited to a linear behavior. To overcome this, Brockwell (1994) extended the class of (Gaussian) CARMA processes by allowing the parameters to be piecewise constant functions depending on the value of the process itself. The resulting processes are called continuous-time threshold autoregressions (CTAR) and allow to successfully map nonlinear relationships. The construction was motivated by Tong (1983), who introduced a similar extension for the class of discrete-time autoregressive processes. While in discrete-time much of the original theory can also be used for the extended process, this is not the case for CTAR. The reason is that in continuous-time

any number of parameter changes between two observations is possible.

Allowing the parameters of a CARMA process to be piecewise constant functions leads to a discontinuous drift coefficient of the multivariate stochastic differential equation (SDE) (1.10). In standard literature such a case is not dealt with. The well-definedness of CTAR have been proved by Brockwell (1994) using the Girsanov theorem. Moreover, a consistent simulation scheme was specified, which forms the basis for a heuristic procedure to estimate model parameters.

2.2 CTAR with jumps

In addition to the occurrence of nonlinear dependencies, it has been found that data often has characteristics that do not interfere with normal distribution assumptions. Especially in financial markets extreme events can often be observed, which suggest the influence of a jump component. While corresponding extensions have been defined for CARMA processes (cf. Brockwell (2001b)), to our knowledge this is not the case for CTAR. Hence, we generalize the model in Brockwell (1994) by incorporating a jump component.

Definition 2.1
A CTAR(p) with jumps $\{X(t)\}_{t\in[0,T]}$, $T > 0$, *of order* $p \in \mathbb{N}$ *is defined as the first component of the p-dimensional process* $\{\boldsymbol{X}(t)\}_{t\in[0,T]}$ *with initial* $x^0 \in \mathbb{R}^p$ *satisfying*

$$d\boldsymbol{X}(t) = \left[A(X(t))\boldsymbol{X}(t) - \boldsymbol{1}_p\beta(X(t))\right]dt + \boldsymbol{1}_p\left[\sigma dW(t) + dJ(t)\right], \tag{2.1}$$

$$\boldsymbol{X}(0) = x^0 \ a.s., \tag{2.2}$$

with $\boldsymbol{1}_p^T = (0 \ \cdots \ 0 \ 1)$, $A(x) = \begin{pmatrix} 0 & I_{p-1} \\ a(x)^T \end{pmatrix}$, *and parameter functions*

$$a(x)^T = (-a_{pi} \ \cdots \ -a_{1i}) \in \mathbb{R}^p, \ \beta(x) = \beta_i \in \mathbb{R}, \ \textit{for } x \in R_i := [r_{i-1}, r_i). \tag{2.3}$$

The threshold values $-\infty = r_0 < r_1 < \cdots < r_l = \infty$ *partition the real line. SDE* (2.1) *is driven by a Lévy jump-diffusion* $L(t) := \sigma W(t) + J(t)$, *where* $\sigma > 0$, $W(t)$ *is a standard Brownian motion and* $J(t) = \sum_{i=1}^{N_t} \gamma_i$ *is a compound Poisson process with jumps of size* $\gamma_i \overset{i.i.d}{\sim} F_\gamma$ *independent of the Poisson process* $N(t)$ *with intensity* $\lambda > 0$. *For* $p = 1$, $A(x)$ *reduces to* $-a_1(x)$.

By (2.3) we clearly see that the drift coefficient of the stochastic differential equation (2.1)

is a discontinuous function. Therefore construction of a (strong) solution by common theorems fail (see e.g. Protter (2005)[V,3.]). Fortunately, by using Girsanov's theorem for semimartingales, a unique weak solution still can be found. A detailed treatment of the theory of semimartingales is given by Jacod and Shiryaev (2002). There, also any unexplained but classical notation used in the following is explained.

Remark 2.2
A weak solution to a SDE driven by a Lévy jump-diffusion $\sigma W + J$ exists, if there is a solution X on the driving system $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}; Z)$, where $Z = \sigma\tilde{W} + \tilde{J}$ is a Lévy jump-diffusion with driving terms $\tilde{W} \stackrel{d}{=} W$, $\tilde{J} \stackrel{d}{=} J$ (see Jacod and Shiryaev (2002)[III,§2c.]).

For application of the Girsanov formula it is essential that for a suitable measurable function $H : [0,T] \times D([0,T],\mathbb{R}) \to \mathbb{R}$, the Doleans-Dade exponential

$$Z(t) := Z(t,X) = \exp\left(\int_0^t H(s,X)dW(s) - \frac{1}{2}\int_0^t H^2(s,X)ds\right),\ 0 \leqslant t \leqslant T, \tag{2.4}$$

is a true martingale. For $X(t) \equiv W(t)$, Beneš approach (see Karatzas and Shreve (1998)[3.5.16]) shows that this is true under a growth condition on H. Klebaner and Lipster (2011) extended this result for more general forms of the stochastic exponential. In the following Lemma we show that under a second order condition on $L(t)$ their results can be applied.

Lemma 2.3
Let $\bar{L}[0,T] := \{L \in L[0,T] : \mathbb{E}(L(t)) = 0,\ \mathbb{E}(L^2(t)) < \infty,\ 0 \leqslant t \leqslant T\}$, where $L[0,T]$ is the class of Lévy processes on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ up to time T. Then for any measurable function $H : [0,T] \times \bar{L}[0,T] \to \mathbb{R}$, where $H(t,L)$ depends on $\{L(s)\}_{s\leqslant t}$ only and satisfying the linear growth condition

$$H^2(t,L) \leqslant K\left[1 + \sup_{s\leqslant t} L^2(s)\right],\ t \leqslant T, \tag{2.5}$$

for some $K > 0$, the Doleans-Dade exponential $\{Z(t)\}_{t\in[0,T]}$ is a martingale.

Proof. First, for the stochastic integral in (2.4) to be well defined, $H(s,L)$ has almost sure to be square integrable with respect to ds on $[0,T]$. Since $\mathbb{E}(L(t)) = 0$, by Cont and Tankov (2004)[3.17], $L(t)$ is a martingale and by Karatzas and Shreve (1998)[1.3.7], $L^2(t)$ is a submartingal. Therefore Doob's inequality can be used to show that by (2.5) and $\mathbb{E}(L^2(t)) < \infty$, $H(t,L)$ satisfies the conditions of Klebaner and Lipster (2011) as

$\bar{L}[0,T] \subset D([0,T],\mathbb{R})$. □

The measurability of the function H in Lemma 2.3 has to hold with respect to an appropriate σ-algebra. In the following we will use the canonical σ-algebra on $D([0,T],\mathbb{R})$, i.e. the Borel σ-algebra associated with the Skorokhod topology.

Using Lemma 2.3 we now can prove that there is a weak solution to (2.1)-(2.2) in sense of Remark 2.2. Furthermore this solution is unique and non-explosive.

Theorem 2.4
For each Lévy jump-diffusion $\{L(t)\}_{t\in[0,T]}$ satisfying the conditions of Lemma 2.3 and $x^0 \in \mathbb{R}^p$, there is a unique (in law), non-explosive weak solution of (2.1)-(2.2).

Proof. First we adopt the idea of Brockwell (1994), attributing a solution of (2.1) to a solution of a one-dimensional SDE. Therefore fix $x^0 \in \mathbb{R}^p$ and w.l.o.g. assume $L(t) = \sigma W(t) + J(t)$ has characteristics $(0, \sigma^2 t, \lambda dt \times F_\gamma)$. Writing equation (2.1) in coordinate form, we get

$$\begin{aligned} dX_1(t) &= X_2(t)dt, \\ dX_2(t) &= X_3(t)dt, \\ &\vdots \\ dX_{p-1}(t) &= X_p(t)dt, \\ dX_p(t) &= [-a_p X_1(t) - \cdots - a_1 X_p(t) - \beta]dt + dL(t), \end{aligned} \tag{2.6}$$

where we have abbreviated $a_j(X_1(t))$ and $\beta(X_1(t))$ to a_j and β, respectively. Assuming $\mathbf{X}(0) = x^0$, we can write $\mathbf{X}(t)$ in terms of $\{X_p(s), 0 \leqslant s \leqslant t\}$ using the relation

$$X_j(t) = x_j^0 + \int_0^t \int_0^{s_{p-1-j}} \cdots \int_0^{s_2} X_p(s_1) ds_1 \cdots ds_{p-j}, \ j = 1, 2, \ldots, p-1.$$

The resulting functional relationship will be denoted by

$$\mathbf{X}(t) = \mathbf{F}(t, X_p). \tag{2.7}$$

Substituting (2.7) into the last equation in (2.6), we see that it can be written in the form

$$dX_p(t) = G(t, X_p)dt + dL(t), \tag{2.8}$$

where $G(t, X_p)$, like $\mathbf{F}(t, X_p)$, depends on $\{X_p(s), 0 \leqslant s \leqslant t\}$.

Let $\Omega = D([0,T],\mathbb{R})$ be the Skorokhod space, $\mathcal{F}$ the canonical σ-algebra and $\mathbb{F} = (\mathcal{F}_t)_{0\leqslant t\leqslant T}$ the canoncial filtration. Now, $G : [0,T] \times D([0,T],\mathbb{R}) \to \mathbb{R}$ is progressively measurable (see Karatzas and Shreve (1998)[1.1.13]).

We now want to show by use of the Girsanov formula for semimartingales, that there is a (weak) solution of Equation (2.8). For this, let $\mathbb{P}$ be a probability measure under which the coordinate mapping process $L(t)(\omega) = \omega(t)$, $0 \leqslant t \leqslant T$, $\omega \in \Omega$, is a Lévy process with the given characteristics. For this mapping $\mathcal{F}^L = \mathcal{F}$. Then we define for all $t \geqslant 0$,

$$Z(t) := Z(t,L) = \exp\left(\int_0^t \frac{G(s,L)}{\sigma} dW(s) - \frac{1}{2}\int_0^t \frac{G^2(s,L)}{\sigma^2} ds\right), \tag{2.9}$$

It is easy to see that by Lemma 2.3, $Z(t)$ is a martingale. Therefore,

$$\tilde{\mathbb{P}}(t,A) := \mathbb{E}(1_A Z(t)),\ A \in \mathcal{F}_t,\ 0 \leqslant t \leqslant T, \tag{2.10}$$

defines a consistent family of probability measures (Karatzas and Shreve (1998)[p. 191]). By Girsanov's theorem for semimartingales (Jacod and Shiryaev (2002)[3.24, III]) the characteristics of $L(t)$ relative to $\tilde{\mathbb{P}}(t)$ are

$$\left(\int_0^t \sigma\rho(s)ds, \sigma^2 t, Y \times t \times \lambda F_\gamma\right), \tag{2.11}$$

where Y is a measurable nonnegative function and $\rho(t)$ is a predictable process uniquely determined by the equations

$$Y = M_{\mu^L}^{\mathbb{P}}(\frac{Z}{Z_-}|\tilde{\mathcal{P}}), \tag{2.12}$$

$$[Z^c, L^c] = \int \sigma\rho(s)Z(s-)ds. \tag{2.13}$$

In Equation (2.12), $M_\mu^{\mathbb{P}}(X|\tilde{\mathcal{P}})$ is the conditional expectation of X on the probability space $(\tilde{\Omega},\tilde{\mathcal{P}},M_\mu^{\mathbb{P}})$ (see Jacod and Shiryaev (2002)[3.15, III] for further details) and $[X,Y] = \{[X,Y](t)\}_{t\in[0,T]}$ is the quadratic covariation of X, Y (see Protter (2005)[6, II]).

As $Z(t)$ is continuous, $Z^c(t) = Z(t-) = Z(t)$. We conclude $Y \equiv 1$ directly and $\rho(s) = \frac{G(s,L(s))}{\sigma}$ since

$$[Z^c, L^c](t) = [Z, \sigma W](t) = [1 + \textstyle\int Z\frac{G}{\sigma} dW, \sigma W](t) =$$

$$\sigma \textstyle\int_0^t Z(s)\frac{G(s,L)}{\sigma} d[W,W](s) = \int_0^t Z(s)G(s,L)ds = \int_0^t \sigma\rho(s)Z(s)ds.$$

Hence by Jacod and Shiryaev (2002)[2.32, II] the process

$$\tilde{L}(t) := L(t) - \int_0^t G(s, L)ds,\ 0 \leqslant t < T, \tag{2.14}$$

has characteristics $(0, \sigma^2 t, \lambda dt \times F_\gamma)$ under $\tilde{\mathbb{P}}$. As $\tilde{L}(t)$ is a Lévy process (see Jacod and Shiryaev (2002)[4,19, II]), $(x^0 + L(t), \tilde{L}(t))$ is a weak solution of (2.8) in sense of Remark 2.2. Furthermore by Klebaner and Lipster (2011)[5.1] this solution does not explode on any finite time interval $[0, T]$, that is $\sup_{t\in[0,T]} \mathbb{E}(X_p(t)^2) < \infty$.

To prove that the solution is unique in law, assume there are two weak solutions (X^i, L^i), $(\Omega^i, \mathcal{F}^i, \mathbb{P}^i)$, $i = 1, 2$, to (2.8) with the same initial value x^0. Now, using the above arguments reversed, the process $X^i(t)$ is a Lévy jump-diffusion with characteristics $(0, \sigma^2 t, \lambda dt \times F_\gamma)$ for the probability measure $\hat{\mathbb{P}}^i(t)$ on $\mathcal{F}_t^i$, according to the prescription $\frac{d\hat{\mathbb{P}}^i(t)}{d\mathbb{P}^i(t)} = \hat{Z}(t, X^i)$, where

$$\hat{Z}(t, X^i) = \exp\left(-\int_0^t \frac{G(s, X^i)}{\sigma} dW(s) - \frac{1}{2}\int_0^t \frac{G^2(s, X^i)}{\sigma^2} ds \right).$$

Therefore, for $0 = t_0 < t_1 < \ldots < t_n \leqslant T$ and $\Gamma \in \mathcal{B}(\mathbb{R}^{n+1})$, we have

$$\mathbb{P}^1\left[(X^1(t_0), \ldots, X^1(t_n)) \in \Gamma\right] = \int_\Gamma \frac{1}{Z(s, X^1)} d\hat{\mathbb{P}}^1 =$$

$$\int_\Gamma \frac{1}{Z(s, X^2)} d\hat{\mathbb{P}}^2 = \mathbb{P}^2\left[(X^2(t_0), \ldots, X^2(t_n)) \in \Gamma\right],$$

concluding that the solution is unique in law. □

2.3 Approximation by a discrete-time process

In practical application it is often necessary to simulate trajectories of the considered stochastic processes. This is because analytic expressions for functional relationships are hard to derive or even do not exist. Therefore Monte Carlo methods are used instead. As for CTAR we do not know how the explicit solution looks like so that, for simulation we have to rely on the defining SDE (2.1) instead. The most common numerical solution to a SDE is given by the Euler method, a first-order approximation.

In the following an Euler representation for the CTAR with jumps is given and we show that this approximation is consistent. This is not trivial as for proving consistency of the Euler approximation one usually requires smoothness of the associated coefficient functions (see e.g. Jacod et al. (2005)).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a solution $\{\mathbf{X}(t)\}_{t\in[0,T]}$ to the SDE (2.1) with

$\mathbf{X}(0) = x^0$ exists. Then an approximation $\{\mathbf{X}^n(\tau_k)\}_{0=\tau_0<\cdots<\tau_n=T,\ \tau_{k+1}-\tau_k =: \delta \equiv T/n}$ to $\{\mathbf{X}(t)\}_{t\in[0,T]}$ is given by

$$\begin{aligned}\mathbf{X}^n(\tau_{k+1}) = \mathbf{X}^n(\tau_k) + \left[A(X^n(\tau_k))\mathbf{X}^n(\tau_k) - \mathbf{1}_p\beta(X^n(\tau_k))\right]\delta + \\ + \mathbf{1}_p\left[\sigma\nu_{k+1}\sqrt{\delta} + \gamma_{k+1}q_{k+1}\right],\ k = 0,\dots,n-1,\end{aligned} \tag{2.15}$$

$$\mathbf{X}^n(\tau_0) = x^0, \tag{2.16}$$

where $\{\nu_k\} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ approximates the increments of a standard Brownian motion, $\{\gamma_k\} \overset{\text{iid}}{\sim} F_\gamma$ is a sequence of stochastic jump amplitudes with zero mean independent of the jump times $\{q_k\} \overset{\text{iid}}{\sim} \text{ber}(\lambda\delta)$. Thereby $\text{ber}(p)$ denotes the Bernoulli distribution with parameter p.

This type of approximation of the process jump dynamic is also used in the Bernoulli diffusion model, a discretized version of the Merton model (see Honoré (1998)). If we extend the discrete time process $\mathbf{X}^n$ to the unit interval by defining $\mathbf{X}^n(t) := \mathbf{X}^n(\lfloor\frac{tn}{n}\rfloor)$, then $\{\mathbf{X}^n(t)\}_{t\in[0,T]}$ is called the discretized Euler scheme.

In the following we prove that (2.15) is a valid approximation of the SDE (2.1) as the distribution of $\mathbf{X}^n$ converges to the distribution of $\mathbf{X}$. The construction of this proof is based on the ideas of Yan (2002), who derived weak consistency of a Euler method with discontinuous coefficients on $C([0,T],\mathbb{R}^p)$ using the occupation time formula. In the multidimensional case this requires the existence of at least one projection of the set of discontinuities onto a coordinate axis with non-degenerate diffusion coefficient such that the projected set has Lebesgue measure 0. Unfortunately, this is not possible in our case. Therefore, beside extending the proof of Yan (2002) to $D([0,T],\mathbb{R}^p)$, we adopt an idea of Brockwell and Williams (1997) to show that the local time technique can still be used.

For this note that $b(\mathbf{X}) := A(X_1)\mathbf{X}$ satisfies a linear growth condition similar to (2.5),

$$\|b(\mathbf{X}^n(t))\|^2 \leqslant K[1 + \|\mathbf{X}^n(t)\|^2],\ 0 \leqslant t \leqslant 1,\ n \in \mathbb{N},$$

for some $K > 0$, where $\|\cdot\|$ stands for the Euclidean norm in the appropriate space. In the following w.l.o.g. we assume that $T = 1$, $\beta = 0$, $x^0 \equiv 0$.

Lemma 2.5
If $m_{\gamma,k} := \mathbb{E}_{F_\gamma}(\gamma_1^k) < \infty$ for $k = 4$, then $\mathbb{E}(\|\boldsymbol{X}^n(t)\|^4) < \infty$ for all $n \in \mathbb{N}$ and $t \in [0,1]$.

Proof. As $\mathbf{X}^n(t) = \mathbf{X}^n(\tau_k)$, $\tau_k \leqslant t < \tau_{k+1}$, we only have to prove $\mathbb{E}(\|\mathbf{X}^n(\tau_k)\|^4) < \infty$ for

$k = 0, 1, \ldots, n$. Let $n \in \mathbb{N}$ be arbitrary but fixed. By

$$\mathbf{X}^n(\tau_{k+1}) = \mathbf{X}_p^n(\tau_k) + \frac{1}{n} b(\mathbf{X}^n(\tau_k)) + \frac{\sigma \nu_{k+1}}{\sqrt{n}} \mathbf{1}_p + \gamma_{k+1} q_{k+1} \mathbf{1}_p,$$

we find

$$\mathbb{E}(\|\mathbf{X}^n(\tau_{k+1})\|^4) \leqslant 4^3 \left[\mathbb{E}(\|\mathbf{X}^n(\tau_k)\|^4) + \frac{2K^2}{n^2} \left(1 + \mathbb{E}(\|\mathbf{X}^n(\tau_k)\|^4)\right) + \frac{3\sigma^4}{n^2} + \frac{\lambda}{n} m_{\gamma,4} \right],$$

where we used $(\sum_{i=1}^m a_i)^4 \leqslant m^3 (\sum_{i=1}^m a_i^2)$ for any real numbers a_i by the Cauchy-Schwarz inequality. Since $x^0 \in \mathbb{R}$, the statement follows now by induction on k. □

Lemma 2.6

If the condition of Lemma 2.5 is satisfied, then $\sup_{n \geqslant 1} \mathbb{E}(\|\boldsymbol{X}^n(t)\|^4) < \infty$ *for all* $t \in [0, 1]$.

Proof. First note that by equation (2.15),

$$\mathbf{X}^n(t) = x^0 + \frac{1}{n} \sum_{i=1}^{\lfloor tn \rfloor} b(\mathbf{X}^n(\tau_{i-1})) + \sum_{i=1}^{\lfloor tn \rfloor} \frac{\sigma \nu_i}{\sqrt{n}} \mathbf{1}_p + \sum_{i=1}^{\lfloor tn \rfloor} \gamma_i q_i \mathbf{1}_p,$$

and so by the Cauchy-Schwarz inequality

$$\|\mathbf{X}^n(t)\|^2 \leqslant 4^3 \left[\left\|x^0\right\|^4 + \left\| \frac{1}{n} \sum_{i=0}^{\lfloor tn \rfloor} b(\mathbf{X}^n(\tau_{i-1})) \right\|^4 + \left(\sum_{i=1}^{\lfloor tn \rfloor} \frac{\sigma \nu_i}{\sqrt{n}} \right)^4 + \left(\sum_{i=1}^{\lfloor tn \rfloor} \gamma_i q_i \right)^4 \right].$$

By the fact that $\sum_{i=0}^{\lfloor tn \rfloor} b(\mathbf{X}^n(\tau_{i-1}))$ is constant for $\frac{\lfloor tn \rfloor}{n} \leqslant t < \frac{\lfloor t(n+1) \rfloor}{n}$, Hölder‘s inequality, the at most linear growth of $\|b(\mathbf{X})\|$ and the Cauchy-Schwarz inequality

$$\begin{aligned}
\|\mathbf{X}^n(t)\|^2 \leqslant & 4^3 \left[\left\|x^0\right\|^4 + t^3 \int_0^t \|b(\mathbf{X}^n(s))\|^4 \, ds + \left(\sum_{i=1}^{\lfloor tn \rfloor} \frac{\sigma \nu_i}{\sqrt{n}} \right)^4 + \left(\sum_{i=1}^{\lfloor tn \rfloor} \gamma_i q_i \right)^4 \right] \\
\leqslant & 4^3 \left[\left\|x^0\right\|^4 + t^3 K \left(t + \int_0^t \|\mathbf{X}^n(s)\|^4 \, ds \right) + \left(\sum_{i=1}^{\lfloor tn \rfloor} \frac{\sigma \nu_i}{\sqrt{n}} \right)^4 + \right. \\
& \left. + \left(\sum_{i=1}^{\lfloor tn \rfloor} \gamma_i q_i \right)^4 \right].
\end{aligned}$$

Therefore there exist two positive constants c_1 and c_2 independent of n such that

$$\mathbb{E}\left(\|\mathbf{X}^n(t)\|^4\right) \leqslant c_1 + c_2 \left[\int_0^t \mathbb{E}\left(\|\mathbf{X}^n(s)\|^4\right) ds \right],$$

where the equality holds by Tonelli's theorem and Lemma 2.5. Let $f_t^n := \mathbb{E}(\|\mathbf{X}^n(s)\|^4)$, then $f_t^n \leqslant c_1 + c_2 \int_0^t f_s^n ds$. Now by Gronwall's lemma, $f_t^n \leqslant c_1 \exp(c_2 t) \leqslant c_1 \exp(c_2)$. This concludes the Lemma, as c_1 and c_2 are independent of n. □

Proposition 2.7
If the condition of Lemma 2.5 is satisfied, then the Euler scheme $\{\boldsymbol{X}^n : n \geqslant 1\}$ *is tight in* $D([0,1], \mathbb{R}^p)$.

Proof. To prove the tightness of the sequence $\{\mathbf{X}^n : n \geqslant 1\}$, we use Ethier and Kurtz (1986)[Theorem 9.8.6]. By applying Markov's inequality and Lemma 2.6, we get first that for all t in $[0,1]$,

$$\limsup_{n \geqslant 1} \mathbb{P}(\|\mathbf{X}^n(t)\| \geqslant a) \leqslant \limsup_{n \geqslant 1} \frac{\mathbb{E}(\|\mathbf{X}^n(t)\|^4)}{a^4} \leqslant \frac{C}{a^4} \underset{a \to \infty}{\to} 0.$$

This shows, that conditions (a) in Ethier and Kurtz (1986)[Theorem 7.2] holds. To use Ethier and Kurtz (1986)[Theorem 9.8.6], by Ethier and Kurtz (1986)[Theorem 9.8.8], it remains to prove that

$$\mathbb{E}\left(\|\mathbf{X}^n(t+h) - \mathbf{X}^n(t)\| \, \|\mathbf{X}^n(t) - \mathbf{X}^n(t-h)\|\right) \leqslant Ch, \ t \in [0,1], \ 0 \leqslant h \leqslant t, \tag{2.17}$$

for some $C > 0$. Note that for $h \geqslant \frac{1}{n}$,

$$\mathbb{E}(\|\mathbf{X}^n(t+h) - \mathbf{X}^n(t)\|^4) = \mathbb{E}\left(\left\|\sum_{i=\lfloor tn \rfloor + 1}^{\lfloor (t+h)n \rfloor} \frac{1}{n} b(\mathbf{X}^n(\tau_{i-1})) + \frac{\sigma \nu_i}{\sqrt{n}} \mathbf{1}_p + \gamma_i q_i \mathbf{1}_p \right\|^4\right) \leqslant$$

$$3^3 \mathbb{E}\left(\left\|\frac{1}{n} \sum_{i=\lfloor tn \rfloor + 1}^{\lfloor (t+h)n \rfloor} b(\mathbf{X}^n(\tau_{i-1}))\right\|^4 + \left(\sum_{i=\lfloor tn \rfloor + 1}^{\lfloor (t+h)n \rfloor} \frac{\sigma \nu_i}{\sqrt{n}}\right)^4 + \left(\sum_{i=\lfloor tn \rfloor + 1}^{\lfloor (t+h)n \rfloor} \gamma_i q_i\right)^4\right) \leqslant$$

$$3^3 h \left(\frac{1}{n} \sum_{i=\lfloor tn \rfloor + 1}^{\lfloor (t+h)n \rfloor} \mathbb{E}(\|b(\mathbf{X}^n(\tau_{i-1}))\|^4) + \frac{3\sigma^2}{n} + \lambda m_{\gamma,4}\right) \leqslant Ch,$$

where $C > 0$ is independent of n and we used the Cauchy-Schwarz inequality, the at most linear growth of $\|b(\mathbf{X})\|$ and Lemma 2.5. Since

$$(\mathbf{X}^n(t+h) - \mathbf{X}^n(t)) \wedge (\mathbf{X}^n(t) - \mathbf{X}^n(t-h)) = 0, \ h < \frac{1}{n},$$

condition (2.17) holds. Therefore $\{\mathbf{X}^n : n \geqslant 1\}$ is tight in $D[0,1]$. □

Since $\{\mathbf{X}^n : n \geqslant 1\}$ is tight in $D([0,1],\mathbb{R}^p)$, which is a separable and complete space under a metric d^0, topologically equivalent to the Skorokhod metric d, by Prohorov's theorem (see Billingsley (1999))[Theorem 5.1], $\{\mathbf{X}^n : n \geqslant 1\}$ is relatively compact in $D[0,1]$. Therefore each sequence $\{\mathbf{X}^{n_i} : i \geqslant 1\}$ contains some subsequence $\{\mathbf{X}^{n_{i(m)}} : m \geqslant 1\}$ converging weakly to some $\mathbf{X}$.

By Skorokhod's representation theorem (see Billingsley (1999)[Theorem 6.7]), there exist random elements $\mathbf{Y}^m$ and $\mathbf{Y}$ taking values in $D([0,1],\mathbb{R}^p)$, defined on a common probability space $(\bar{\Omega},\bar{\mathcal{F}},\bar{\mathbb{P}})$, such that $\mathcal{L}(\mathbf{Y}^m) = \mathcal{L}(\mathbf{X}^{n_{i(m)}})$, $\forall m$, $\mathcal{L}(\mathbf{Y}) = \mathcal{L}(\mathbf{X})$ and $\mathbf{Y}^m \underset{m\to\infty}{\to} \mathbf{Y}$ almost surely in $D([0,1],\mathbb{R}^p)$. In addition, by Van der Vaart and Wellner (1996)[Addendum 1.10.5], $\mathbf{Y}^m$ and $\mathbf{Y}$ can be chosen according to

$$\mathbf{Y}^m(\cdot,\bar{\omega}) = \mathbf{X}^{n_{i(m)}}(\cdot,\phi^m(\bar{\omega})),\ \mathbf{Y}(\cdot,\bar{\omega}) = \mathbf{X}(\cdot,\phi(\bar{\omega})), \tag{2.18}$$

with measurable maps $\phi^m : \bar{\Omega} \to \Omega$ and $\mathbb{P} = \bar{\mathbb{P}} \circ \phi^m$, for $m = 1,2,\ldots$.

If we define $\bar{\nu}_k^m := \nu_k \circ \phi^m$, $\bar{\gamma}_k^m := \gamma_k \circ \phi^m$ and $\bar{q}_k^m := q_k^m \circ \phi^m$, $m = 1,2,\ldots$, the distribution of the random variables is not changed under $\bar{\mathbb{P}}$. Therefore for every $m \geqslant 1$, $\mathbf{Y}^m$ satisfies

$$\mathbf{Y}^m(t) = \mathbf{Y}^m(\tau_k) + \frac{1}{n_{i(m)}} A(Y_1^m(\tau_k))\mathbf{Y}^m(\tau_k) + \mathbf{1}_p \left[\frac{\sigma \bar{\nu}_{k+1}^m}{\sqrt{n_{i(m)}}} + \bar{\gamma}_{k+1}^m \bar{q}_{k+1}^m \right],\ \tau_k \leqslant t < \tau_{k+1}. \tag{2.19}$$

By this representation of $\mathbf{Y}^m$, Yan (2002) proved, that $\mathbf{Y}$ is a weak solution of the approximated SDE. Hence $\{\mathbf{X}^{n_{i(m)}} : m \geqslant 1\}$ converges weakly to the unique weak solution, which implies weak convergence of the Euler scheme. By the local time technique of Yan (2002), we now show that the occupation time of Y_1 in the set of discontinuities $D_a := \{x \in \mathbb{R} : x = r_i,\ i = 1,\ldots,l\}$ of the drift function b is of Lebesgue measure 0 almost surely. Let $C_b^2(\mathbb{R})$ denote the space of continuous functions f on $\mathbb{R}$ with continuous first two derivatives bounded by a constant $b > 0$.

Lemma 2.8

If the condition of Lemma 2.5 is satisfied, then for $t \in [0,1]$,

$$[Y_p](t) := [Y_p, Y_p](t) = t(\sigma^2 + m_{J,2}\lambda),\ \ a.s..$$

Furthermore, for any $f \in C_b^2(\mathbb{R})$,

$$[f(Y_i), Y_p](t) = 0,\ i \neq p,\ \ a.s..$$

Proof. We first prove, that

$$[Y_p^m](t) \overset{L^1(\Omega)}{\underset{m\to\infty}{\to}} \int_0^t (\sigma^2 + m_{J,2}\lambda)ds. \tag{2.20}$$

Note, by the almost sure representation (2.19),

$$Y_p^m(\cdot) = y_p^0 + \frac{1}{n_{i(m)}} \sum_{k=1}^{\lfloor \cdot n_{i(m)} \rfloor} b(\mathbf{Y}^m(\tau_{k-1}))\mathbf{1}_p + \sum_{k=1}^{\lfloor \cdot n_{i(m)} \rfloor} \frac{\sigma\bar{\nu}_k^m}{\sqrt{n_{i(m)}}} + \sum_{k=1}^{\lfloor \cdot n_{i(m)} \rfloor} \bar{\gamma}_k^m \bar{q}_k^m.$$

$Y_p^m(t,\bar{\omega})$ is a cadlag function in t and therefore bounded on $[0,1]$ for each $\bar{\omega} \in \bar{\Omega}$. As $Y_p^m(t,\bar{\omega})$ is constant in t everywhere except of its $n_{i(m)}$ jump points, it is of finite variation almost sure. Therefore $Y_p^m(t)$ is a pure jump semimartingale with respect to its natural filtration $\bar{\mathcal{F}}_t^m = \sigma(Y_p^m(t) : 0 \leqslant t \leqslant 1)$.

Thus, by Jacod and Shiryaev (2002)[I.4.52],

$$[Y_p^m](t) = \sum_{k=0}^{\lfloor tn_{i(m)} \rfloor} (\Delta Y_p^m(\tau_k))^2 = \sum_{k=1}^{\lfloor tn_{i(m)} \rfloor} \left(\frac{1}{n_{i(m)}} b(\mathbf{Y}^m(\tau_{k-1}))\mathbf{1}_p + \frac{\sigma\bar{\nu}_k^m}{\sqrt{n_{i(m)}}} + \bar{\gamma}_k^m \bar{q}_k^m \right)^2 =$$
$$\sum_{k=1}^{\lfloor tn_{i(m)} \rfloor} \left[\frac{1}{n_{i(m)}^2} b(\mathbf{Y}^m(\tau_{k-1}))^2\mathbf{1}_p + \frac{\sigma^2(\bar{\nu}_k^m)^2}{n_{i(m)}} + (\bar{\gamma}_k^m)^2(\bar{q}_k^m)^2 + \right.$$
$$\left. + \frac{2\sigma\bar{\nu}_k^m}{n_{i(m)}^{3/2}} b(\mathbf{Y}^m(\tau_{k-1}))\mathbf{1}_p + \frac{2\bar{\gamma}_k^m\bar{q}_k^m}{n_{i(m)}} b(\mathbf{Y}^m(\tau_{k-1}))\mathbf{1}_p + \frac{2\sigma\bar{\nu}_k^m}{\sqrt{n_{i(m)}}} \bar{\gamma}_k^m \bar{q}_k^m \right].$$

and so

$$\mathbb{E}[Y_p^m](t) = \sum_{k=1}^{\lfloor tn_{i(m)} \rfloor} \left[\frac{1}{n_{i(m)}^2} \mathbb{E}\left(b(\mathbf{Y}^m(\tau_{k-1}))^2\mathbf{1}_p \right) + \frac{\sigma^2}{n_{i(m)}} + \frac{m_{\gamma,2}\lambda}{n_{i(m)}} \right],$$

where $\Delta X(t) = X(t) - X(t-)$. As $\|b(\mathbf{X})\|$ has at most linear growth and Lemma 2.5, statement (2.20) follows. At the same time, by Kurtz and Protter (1991)[Theorem 2.2], $[Y_p^m](t) \underset{m\to\infty}{\overset{p}{\to}} [Y_p](t)$, as $[X] = X^2 - 2\int X_- dX$ and $\mathbf{Y}^m \underset{m\to\infty}{\to} \mathbf{Y}$ a.s..

From Lemma 2.5 it can be seen in the same way as above that $\mathbb{E}[Y_p^m]^2(t)$ is uniformly bounded for all t and m. Therefore $[Y_p^m]$ is uniformly integrable, from which L^1 convergence follows. This proves the first statement of the Lemma.

For the proof of the second statement w.l.o.g. assume $p = 2$. Note that $f(Y_p^m)$ is still a semimartingale as f is twice continuously differentiable. Hence, by the Lipschitz conti-

nuity of f and the Cauchy-Schwartz inequality, we get

$$\mathbb{E}[f(Y_1^m), Y_2^m]^2(t) = \mathbb{E}\left(\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \Delta f(Y_1^m(\tau_k))\Delta Y_2^m(\tau_k)\right)^2 \leqslant$$

$$\lfloor tn_{i(m)}\rfloor \mathbb{E}\left(\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} (\Delta f(Y_1^m(\tau_k)))^2(\Delta Y_2^m(\tau_k))^2\right) \leqslant$$

$$\lfloor tn_{i(m)}\rfloor b^2 \mathbb{E}\left(\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} (\Delta (Y_1^m(\tau_k)))^2(\Delta Y_2^m(\tau_k))^2\right) \leqslant$$

$$\frac{tb^2}{n_{i(m)}} \sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \left(\mathbb{E}(Y_2^m(\tau_{k-1})^4)\right)^{\frac{1}{2}} \left(\mathbb{E}(\Delta Y_2^m(\tau_k))^4\right)^{\frac{1}{2}} \leqslant \frac{Cb^2}{n_{i(m)}} \underset{m\to\infty}{\to} 0,$$

where for the last inequality we used Lemma 2.5 and $\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} (\mathbb{E}(\Delta Y_2^m(\tau_k))^4)^{\frac{1}{2}} \leqslant C$, for some $C > 0$ independent of $n_{i(m)}$. This can be proved in the same way as in the proof of Proposition 2.7.

By Kurtz and Protter (1991)[Theorem 2.2], we know in addition that $[f(Y_1^m), Y_2^m](t) \underset{m\to\infty}{\overset{p}{\to}} [f(Y_1), Y_2](t)$, as $[X, Y] = XY - \int X_- dY - \int Y_- dX$, if $(f(Y_1^m), Y_2^m) \underset{m\to\infty}{\overset{p}{\to}} (f(Y_1), Y_2))$ in the Skorokhod topology on $D([0, 1], \mathbb{R}^2)$. But this is obvious, as $\mathbf{Y}^m \underset{m\to\infty}{\overset{\text{a.s.}}{\to}} \mathbf{Y}$ and the continuous mapping theorem. Therefore, by the same argument as above, $[f(Y_1), Y_2](t) = 0$, $\forall t$, a.s.. □

Lemma 2.9

If the condition of Lemma 2.5 is satisfied, then

$$\int_0^1 \mathbf{1}(Y_1(s) \in D_a)ds = 0, \ \ a.s..$$

Proof. We first consider the case $p = 1$. In this setting the conditions in Yan (2002) are satisfied and so we simply can use the occupation time formula for semimartingales (see Protter (2005)[Ch. IV, Corollary 1]). This is valid, since by Kurtz and Protter (1991)[Theorem 2.2], Y_1 is a semimartingale. Let L^x be the local time of Y_1 at x and Y^c the continuous part of Y_1. Then by Lemma 2.8,

$$\int_0^1 \mathbf{1}(Y_1^c(s) \in D_a)(\sigma^2 + m_{J,2}\lambda)ds = \int_0^1 \mathbf{1}(Y_1^c(s) \in D_b)d[Y_1^c](s) = \int_{x\in D_b} L_1^x dx = 0,$$

where we used $[Y_1](t) = [Y_1]^c(t)$ by Lemma 2.8 and $[Y_1]^c(t) = [Y_1^c](t)$ (see Protter

(2005)[p. 63]). The last equation holds as $\lambda(D_b) = 0$ and $L^x(t) < \infty$ a.s.. As each cadlag function has at most countable many discontinuities (Billingsley (1999)[p. 124]), we conclude the Lemma by $\sigma^2 + m_{J,2}\lambda > 0$.

To prove the Lemma in case of $p \geqslant 2$ we adopt an argument of Brockwell and Williams (1997), to show that even for a degenerate diffusion coefficient the amount of time that Y_1 spends in a neighborhood of r_1 is small with respect to the Lebesgue measure. W.l.o.g. we only consider $p = 2$, $l = 2$ and $r_1 = 0$.

For $K > 0$ define $\nu_K := \inf\{t \in [0,1] : \|\mathbf{Y}(t)\| \geqslant K\}$ and fix ϵ such that $K > 1 > 2\epsilon > 0$. Then we choose a function $g \in C_b^3(\mathbb{R})$, where g is such that $g(y) = \int_0^y \int_0^w g''(u)dudw$ and g'' is an even function, $g''(y) = 1$ for $0 \leqslant y \leqslant \epsilon$, $g''(y) = 0$ for $2\epsilon \leqslant y \leqslant K$ and $g''(y)$ is non-increasing for $\epsilon \leqslant y \leqslant 2\epsilon$. In particular for $|y| \leqslant K$, $|g'(y)| \leqslant 2\epsilon$ and $|g(y)| \leqslant 2\epsilon K$. Thus, we may assume g'' is defined for $|y| > K$ such that $|g(y)| \leqslant 4\epsilon K$ and $|g'(y)| \leqslant 2\epsilon$ for all y.

Brockwell and Williams (1997) used a martingale property of $Y_2 \times g'(Y_1)$ to show $\lambda(\{|Y_1| \leqslant \epsilon\}) \underset{\epsilon \to 0}{\to} 0$. Instead, we utilize a more direct approach based on the integration by parts formula for semimartingales (see Protter (2005)[II.2]) to derive

$$\mathbb{E}\left(\int_0^{t\wedge\nu_K} g''(Y_1(s))Y_2^2(s)ds\right) = \mathbb{E}\left(Y_2(t\wedge\nu_K)g'(Y_1(t\wedge\nu_K))\right) -$$

$$\mathbb{E}\left(\int_0^{t\wedge\nu_K} g'(Y_1(s))dY_2(s)\right) - [g'(Y_1), Y_2](t\wedge\nu_K) \leqslant CKt\epsilon, \text{ a.s.},$$

by Lemma 2.6, Lemma 2.8 and the boundedness of g', where $C > 0$. Thus, on letting ϵ tend to zero, $t \to 1$, and then $K \to \infty$,

$$\mathbb{E}\left(\int_0^1 Y_2^2(s)\mathbf{1}(0 = Y_1(s))ds\right) = 0.$$

As $\int_0^1 \mathbf{1}(0 = Y_2(s))ds = 0$, a.s. (see case $p = 1$), it follows $\int_0^1 \mathbf{1}(0 = Y_1(s))ds = 0$, a.s.. This completes the proof. □

Lemma 2.10

If the conditions of Lemma 2.5 are satisfied, then

$$\frac{1}{n_{i(m)}} \sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} A(Y_1^m(\tau_{k-1}))\,\boldsymbol{Y}^m(\tau_{k-1}) \overset{L^1(\Omega)}{\underset{m\to\infty}{\to}} \int_0^t A(Y_1(s))\,\boldsymbol{Y}(s)ds.$$

Proof. As the first $p-1$ components of $A(X_1)\mathbf{X}$ are continuous functions of $\mathbf{X}$ by

$\mathbf{Y}^m \overset{\text{a.s.}}{\underset{m\to\infty}{\to}} \mathbf{Y}$, Lemma 2.6 and the dominated convergence theorem, the statement can be proved directly for all components unequal to p. For the last component we have to use Lemma 2.9 in addition.

By Skorokhod's representation theorem,

$$d(\mathbf{Y}^m, \mathbf{Y}) \underset{m\to\infty}{\to} 0, \text{ a.s..}$$

As Skorokohod convergence implies that $\mathbf{Y}^m(t) \underset{m\to\infty}{\to} \mathbf{Y}(t)$ for all continuity points t of $\mathbf{Y}$ (Billingsley (1999)[p. 124]), we have

$$b_p(\mathbf{Y}^m(t)) \underset{m\to\infty}{\to} b_p(\mathbf{Y}(t)), \ \forall t \in \overline{D_{\mathbf{Y}}} \cap \overline{D_{a(Y_1)}}, \text{ a.s.},$$

where $D_{\mathbf{Y}}$ is the set of jump times of $\mathbf{Y}$ and $D_{a(Y_1)} := \{t \in [0,1] : Y_1 \in D_a\}$. As $\mathcal{L}(Y^m) = \mathcal{L}(X^{n_{i(m)}})$, by the at most linear growth of $\|b(\mathbf{X})\|$ and the proof of Lemma 2.6, $\sup_{m\geqslant 1} \mathbb{E}\,\|b(\mathbf{Y}^m(t))\|^4$ is dominated by some constant on $[0,1]$. Therefore $\{b_p(\mathbf{Y}^m(t))\mathbf{1}_{\{t\in\overline{D_{\mathbf{Y}}}\cap\overline{D_{a(Y_1)}}\}}\}_{m\geqslant 1}$ is uniformly integrable. By Vitali's convergence theorem,

$$b_p(\mathbf{Y}^m(t))\mathbf{1}_{\{t\in\overline{D_{\mathbf{Y}}}\cap\overline{D_{a(Y_1)}}\}} \overset{L^1(\Omega)}{\underset{m\to\infty}{\to}} b_p(\mathbf{Y}(t))\mathbf{1}_{\{t\in\overline{D_{\mathbf{Y}}}\cap\overline{D_{a(Y_1)}}\}}, \ \forall t \in [0,1]. \tag{2.21}$$

Hence, we get

$$\mathbb{E}\left|\frac{1}{n_{i(m)}}\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} b_p(\mathbf{Y}^m(\tau_{k-1})) - \int_0^t b_p(\mathbf{Y}(s))ds\right| \leqslant$$
$$\mathbb{E}\left|\frac{1}{n_{i(m)}}\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} b_p(\mathbf{Y}^m(\tau_{k-1})) - \int_0^t b_p(\mathbf{Y}^m(s))ds\right| + \mathbb{E}\left|\int_0^t b_p(\mathbf{Y}^m(s))ds - \int_0^t b_p(\mathbf{Y}(s))ds\right| \leqslant$$
$$\mathbb{E}|b_p(\mathbf{Y}^m_{\tau_{\lfloor tn_{i(m)}\rfloor-1}})||t-\tau_{\lfloor tn_{i(m)}\rfloor-1}| + \int_0^t \mathbb{E}|b_p(\mathbf{Y}^m(s)) - b_p(\mathbf{Y}(s))|ds \underset{m\to\infty}{\to} 0,$$

since $|t-\tau_{\lfloor tn_{i(m)}\rfloor-1}| \leqslant \frac{1}{n_{i(m)}}$, $\lambda\left(D_{\mathbf{Y}} \cup D_{a(Y_1)}\right) = 0$ by Lemma 2.9, equation (2.21) and the dominated convergence theorem. □

Theorem 2.11

If $\mathbb{E}_{F_\gamma}(\gamma_1^4) < \infty$, then the Euler scheme defined in (2.15) *weakly converges to the unique weak solution of SDE* (2.1) *as $n \to \infty$.*

Proof. We define

$$\begin{aligned}\mathbf{Z}^m(t) &:= \mathbf{Y}^m(t) - \frac{1}{n_{i(m)}} \sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} A(Y_1^m(\tau_{k-1}))\mathbf{Y}^m(\tau_{k-1}) \\ &= \mathbf{1}_p \left[\sum_{k=1}^{\lfloor tn_{i(m)}\rfloor} \frac{\sigma \bar{\nu_k}^m}{\sqrt{n_{i(m)}}} + \sum_{K=1}^{\lfloor tn_{i(m)}\rfloor} \bar{\gamma_k}^m \bar{q_k}^m \right],\end{aligned}$$

$$\mathbf{Z}(t) := \mathbf{Y}(t) - \int_0^t A(Y_1^m(s))\mathbf{Y}^m(s)ds.$$

As $\mathbf{Y}^m \underset{m\to\infty}{\to} \mathbf{Y}$ a.s. and Lemma 2.10, $\mathbf{Z}^m$ converges to $\mathbf{Z}$ in probability. Therefore, we only have to prove

$$Z_p^m(t) \underset{m\to\infty}{\overset{d}{\to}} \sigma W(t) + \sum_{i=1}^{N(t)} \gamma_i =: Z^*(t). \tag{2.22}$$

Then, $Y_p(t) \overset{d}{=} \int_0^t b_p(\mathbf{Y}(s))ds + \sigma W(t) + \sum_{i=1}^{N(t)} \gamma_i$, that is, $\mathbf{Y}$ is the unique weak solution of SDE (2.1). Since $\mathbf{X}^{n_{i(m)}}$ converges weakly to $\mathbf{Y}$ and $\mathcal{L}(\mathbf{Y}) = \mathcal{L}(\mathbf{X})$, the Euler scheme converges weakly to the unique weak solution of SDE (2.1).

To prove (2.22), we use Billingsley (1999)[Theorem 13.5]. First, for any $\epsilon > 0$,

$$\mathbb{P}(|Z^*(1) - Z^*(1-)| > \epsilon) = \mathbb{P}(|\gamma_{N(1)}(N(1) - N(1-\delta))| > \epsilon) = 0,$$

by the continuity in probability of $N(t)$. Next, denote by $T_N \subset [0,1]$ the set of discontinuities of $N(t)$. W.l.o.g. let $t_1 < t_2$, $t_1, t_2 \in T_N$. Then

$$\mathbb{E}e^{i(s_1 Z^m(t_1) + s_2 Z^m(t_2))} = \mathbb{E}e^{i(s_1+s2)Z^m(t_1) + is_2(Z^m(t_2) - Z^m(t_1))} =$$

$$\mathbb{E}e^{i(s_1+s2)\sum_{k=1}^{\lfloor t_1 n_{i(m)}\rfloor} \frac{\sigma\bar{\nu}_k^m}{\sqrt{n_{i(m)}}} + is_2 \sum_{k=\lfloor t_1 n_{i(m)}\rfloor+1}^{\lfloor t_2 n_{i(m)}\rfloor} \frac{\sigma\bar{\nu}_k^m}{\sqrt{n_{i(m)}}}} \times$$

$$\mathbb{E}e^{i(s_1+s2)\sum_{k=1}^{\lfloor t_1 n_{i(m)}\rfloor} \gamma_k^m q_k^m + is_2 \sum_{k=\lfloor t_1 n_{i(m)}\rfloor+1}^{\lfloor t_2 n_{i(m)}\rfloor} \gamma_k^m q_k^m}.$$

By Donsker's theorem (see Billingsley (1999)[Theorem 14.1]), the first factor converges to $\mathbb{E}e^{i(s_1 W(t_1) + s_2 W(t_2))}$. To prove that the second factor tends to $\mathbb{E}e^{i(s_1 \sum_{i=1}^{N(t_1)} \gamma_i + s_2 \sum_{i=1}^{N(t_2)} \gamma_i)}$, we show $\sum_{i=1}^{\lfloor tn\rfloor} \bar{\gamma}_i^m \bar{q}_i^m \underset{n\to\infty}{\overset{d}{\to}} \sum_{i=1}^{N(t)} \gamma_i$. This is true, because by the law of total probability

$$\mathbb{P}(\sum_{i=1}^{\lfloor tn\rfloor} \bar{\gamma}_i^m \bar{q}_i^m \leqslant x) = \sum_{k=1}^{\infty} F_{\gamma_1+\ldots+\gamma_k}(x)\mathrm{bin}(\lfloor tn\rfloor, \frac{\lambda}{n})(k),$$

where $\text{bin}(\lfloor tn \rfloor, \frac{\lambda}{n}) \underset{n\to\infty}{\to} \text{pois}(\lambda t)$, since $\frac{\lfloor tn \rfloor}{n}\lambda \underset{n\to\infty}{\to} \lambda t$. Now, to conclude the proof of (2.22), it is sufficient that Z^m fulfills a certain tightness condition, namely for $s \leqslant u \leqslant t$, $s, u, t \in [0,1]$, and $m \geqslant 1$,

$$\mathbb{E}|Z^m(u) - Z^m(s)|^{2\beta}|Z^m(t) - Z^m(u)|^{2\beta} \leqslant c(t-s)^{2\alpha},$$

where $\beta \geqslant 0$, $\alpha > 1/2$ and $c > 0$. This is true, as

$$\mathbb{E}|Z^m(u) - Z^m(s)|^2|Z^m(t) - Z^m(u)|^2 = \mathbb{E}|Z^m(u) - Z^m(s)|^2\mathbb{E}|Z^m(t) - Z^m(u)|^2 \leqslant$$

$$\leqslant \left(\mathbb{E}\left| \sum_{k=\lfloor sn_{i(m)} \rfloor + 1}^{\lfloor un_{i(m)} \rfloor} \frac{\sigma \bar{\nu}_k^m}{\sqrt{n}} \right|^2 + \mathbb{E}\left| \sum_{k=\lfloor sn_{i(m)} \rfloor + 1}^{\lfloor un_{i(m)} \rfloor} \bar{\gamma}^m \bar{p}_k^m \right|^2 \right) \times$$

$$\left(\mathbb{E}\left| \sum_{k=\lfloor un_{i(m)} \rfloor + 1}^{\lfloor tn_{i(m)} \rfloor} \frac{\sigma \bar{\nu}_k^m}{\sqrt{n}} \right|^2 + \mathbb{E}\left| \sum_{k=\lfloor un_{i(m)} \rfloor + 1}^{\lfloor tn_{i(m)} \rfloor} \bar{\gamma}^m \bar{p}_k^m \right|^2 \right) \leqslant$$

$$\leqslant (\lfloor un_{i(m)} \rfloor - \lfloor sn_{i(m)} \rfloor)(\lfloor tn_{i(m)} \rfloor - \lfloor un_{i(m)} \rfloor) \left(\frac{\sigma^2}{n} + m_{J,2}\frac{\lambda}{n}(1 - \frac{\lambda}{n}) \right)^2 \leqslant$$

$$\leqslant \begin{cases} c(u - s + \frac{1}{n})(t - u + \frac{1}{n}) \leqslant 4c(t-s)^2, & t - s \geqslant \frac{1}{n}, s < u < t, \\ 0 & \text{else.} \end{cases}$$

□

2.4 Statistical inference

In this section the problem of fitting a CTAR with jumps to a finite set of possibly irregularly spaced data is considered. For (linear) CAR the explicit solution to the SDE (2.1) allows to calculate the Gaussian likelihood of the observations with help of the discrete-time Kalman recursions (cf. Section 1.5). For CTAR an explicit solution is not found and so this approach cannot be used.

If the data additionally is uniformly spaced, an alternative procedure for estimation of CAR processes is given by their discrete-time representation. A sampled CAR process satisfies a standard ARMA equation. By this fact fitting of a CAR process can be traced back to fitting of an ARMA process (see Brockwell (2014) and references therein). Indeed, such a relationship does not hold for CTAR as noted by Hyndman (1992).

Nonetheless, Gaussian CTAR models have been fitted to a variety of data sets. If the data is observed frequently one can use the stochastic exponential (2.4) to get estimators

of the autoregressive parameters (see Brockwell et al. (2007)). This approach is also possible for the CTAR with jumps but since high-frequency data is rarely available in practice, we do not want to go into detail here. Instead, we want to introduce an approach that can be used in a very general setup. This approach is based on particle filtering methods.

Let $\mathbf{Y}_n := \{y(t_1), y(t_2), \ldots, y(t_n)\}$, $t_1 < t_2 < \ldots < t_n$, be a set of observations, where w.l.o.g. we assume that $t_i - t_{i-1} = \delta \ \forall i$ as the introduced procedures extend in an obvious way for irregularly spaced data. For fitting of a CTAR process we are interested in evaluation of the likelihood $L(\theta; \mathbf{Y}_n)$, being given by

$$L(\theta; \mathbf{Y}_n) = \pi_\theta(y(t_1)) \prod_{i=2}^{n} \pi_\theta(y(t_i)|\mathbf{Y}_{i-1}), \tag{2.23}$$

where π_θ denotes a density specified by the vector of all model parameters θ.

A problem arising in representation (2.23) is that the densities $\pi_\theta(y(t_i)|\mathbf{Y}_{i-1})$ for the CTAR process are unknown. This makes a direct implementation of a maximum likelihood approach impossible. A possible way out is to estimate the parameters based on an estimator of the likelihood itself. A method to find a suitable estimate $\widehat{L}(\theta; \mathbf{Y}_n)$ is given via *particle filters* also known as *sequential monte carlo* methods (see Pitt (2002)).

Particle filters are Monte Carlo type algorithms that represent the posterior density of a stochastic process by sampling a set of particles. They are designed for hidden Markov models, where the observations $\{y_t|\mathbf{X}_t\}$, conditional on a preliminary assumed to be unknown state $\mathbf{X}_t$, being independent and $\{\mathbf{X}_t\}$ is assumed to be Markovian. Obviously for our observations $y(t_i)$ and the state vector $\mathbf{X}(t_i)$ of (2.1) these assumptions are satisfied. Pitt (2002) used the representation

$$\pi_\theta(y(t_i)|\mathbf{Y}_{i-1}) = \int g_\theta(y(t_i)|\mathbf{X}(t_i)) f_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1}) d\mathbf{X}(t_i), \tag{2.24}$$

to estimate the densities with help of Monte Carlo integration. Here g_θ is called the observation density and f_θ the transition density. As for the CTAR process

$$g_\theta(y(t)|\mathbf{X}(t)) = \delta_{X_1(t)}(y(t)), \tag{2.25}$$

direct application of (2.24) is infeasible. Instead we restrict to the unknown state vector

$\bar{\mathbf{X}} := (X_2, \ldots, X_p)$, and use

$$\pi_\theta(y(t_i)|\mathbf{Y}_{i-1}) = \int g_\theta(y(t_i)|\bar{\mathbf{X}}(t_{i-1}), y(t_{i-1}))f_\theta(\bar{\mathbf{X}}(t_{i-1})|\mathbf{Y}_{i-1})d\bar{\mathbf{X}}(t_{i-1}). \tag{2.26}$$

instead. Now the integrand is not longer known explicitly, but can be approximated by

$$g_\theta(y(t_i)|\bar{\mathbf{X}}^k(t_{i-1}), \mathbf{Y}_{i-1}) =$$

$$\int f_\theta(y(t_i), \bar{\mathbf{X}}(t_i)|\bar{\mathbf{X}}^k(t_{i-1}), y(t_{i-1}))d\bar{\mathbf{X}}(t_i) \approx$$

$$\int \frac{1}{L}\sum_{j=1}^{L} K_h(y(t_i) - y^{kj}(t_i))K_h(\bar{\mathbf{X}}(t_i) - \bar{\mathbf{X}}^{kj}(t_i))d\bar{\mathbf{X}}(t_i) =$$

$$\frac{1}{L}\sum_{j=1}^{L} K_h(y(t_i) - y^{kj}(t_i)), \tag{2.27}$$

where K_h is a kernel with bandwidth h and $\{(y, \bar{\mathbf{X}})^{kj}(t_i)\}_{1\leqslant j\leqslant L}$ are random variables from $f_\theta(\mathbf{X}(t_i)|\bar{\mathbf{X}}^k(t_{i-1}), y(t_{i-1}))$, where $\bar{\mathbf{X}}^k(t_{i-1}) \sim \pi_\theta(\bar{\mathbf{X}}(t_{i-1})|\mathbf{Y}_{i-1})$. π_θ is also known as the filtering density. Now, by Monte Carlo integration of (2.26), we get

$$\hat{\pi}_\theta(y(t_i)|\mathbf{Y}_{i-1}) = \frac{1}{NL}\sum_{k=1}^{N}\sum_{j=1}^{L} K_h(y(t_i) - y^{kj}(t_i)). \tag{2.28}$$

By the properties of kernel density estimators, convergence of $\hat{\pi}_\theta(y(t_i)|\mathbf{Y}_{i-1})$ to $\pi_\theta(y(t_i)|\mathbf{Y}_{i-1})$ is ensured. But this approximation requires $N \times L$ random number generations.

A more efficient approximation scheme is given by the *convolution particle filter* of Rossi and Vila (2006). Rossi and Vila (2006) proposed to sample from the joint density $\pi(y(t_i), \bar{\mathbf{X}}(t_i)|\mathbf{Y}_{i-1}))$ by first simulating $\mathbf{X}^k(t_i) \sim f_\theta(\mathbf{X}(t_i)|\bar{\mathbf{X}}^k(t_{i-1}), y(t_{i-1}))$ followed by generation of an observation according to $g_\theta(y(t_i)|\mathbf{X}^k(t_i))$. Then the density of $(y(t_i), \mathbf{X}(t_i))$ is approximated by $(y^k(t_i), \mathbf{X}^k(t_i))$ with help of kernel density estimation. For CTAR by (2.25) we only have to simulate from $f_\theta(\mathbf{X}(t_i)|\bar{\mathbf{X}}^k(t_{i-1}), y(t_{i-1}))$ which can be done directly with the Euler scheme presented in Section 2.3. Then

$$\pi_\theta(\bar{\mathbf{X}}(t_i)|\mathbf{Y}_i) = \frac{\pi_\theta(\bar{\mathbf{X}}(t_i), y(t_i)|\mathbf{Y}_{i-1})}{\pi_\theta(y(t_i)|\mathbf{Y}_{i-1})} \approx$$

$$\frac{\frac{1}{N}\sum_{k=1}^{N} K_h(y(t_i) - y^k(t_i))K_h(\bar{\mathbf{X}}(t_i) - \bar{\mathbf{X}}^k(t_i))}{\frac{1}{N}\sum_{k=1}^{N} K_h(y(t_i) - y^k(t_i))}. \tag{2.29}$$

This approximation allows to generate samples according to $\pi_\theta(y(t_i)|\mathbf{Y}_{i-1})$. For that, starting with a guess of the initial states, we recursively resample from the set $\{\bar{\mathbf{X}}^k(t_i)\}_k$ with weights proportional to (2.29) followed by a Euler step. Thus, an estimator for the likelihood of the CTAR with jumps is

$$\widehat{L}(\theta, \mathbf{Y}_n) = \hat{\pi}_\theta(y(t_1)) \prod_{i=2}^{n} \hat{\pi}_\theta(y(t_i)|\mathbf{Y}_{i-1}), \tag{2.30}$$

where $L = 1$ in (2.28).

Remark 2.12
Brockwell (1994) replaced the transition densities in (2.26) *by Gaussian densities with fitted first and second order moments and used Riemann sums to approximate the integrals. Beside the lack of theoretical backing, for the highly nonlinear CTAR with jumps model this method can be very inaccurate. Moreover, for large p and n, this approach suffers from the complicated approximation of higher-dimensional integrals.*

We now try to find the maximum likelihood estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ L(\theta, \mathbf{Y}_n),$$

by replacing the true likelihood $L(\theta, \mathbf{Y}_n)$ by its approximation $\widehat{L}(\theta, \mathbf{Y}_n)$. Sampling with respect to the the true transition density, Vila (2012) proved that (2.29) and (2.30) converge almost surely to the true filtering density resp. likelihood by convergence properties of kernel density estimators. A similar result is still true if we use the Euler approximation at the evolving step.

Therefore we assume K to be a bounded, positive, symmetrical function from $\mathbb{R}^p \to \mathbb{R}$, such that $\int K d\lambda = 1$, where λ is the Lebesgue measure. If we further assume $\lim_{||x||\to\infty} ||x|| K(x) = 0$, K_h is called *Parzen-Rosenblatt kernel*.

Lemma 2.13
Let $K_{h_N}(x) := \frac{1}{h_N^p} K(\frac{x}{h_N})$, *where* K *is a Parzen-Rosenblatt kernel and* h_N *the bandwidth parameter with* $0 < h_N \searrow 0$ *and* $N h_N^p \to \infty$ *as* $N \to \infty$. *Furthermore we assume the transition density* f_θ *is a positive function satisfying* $f_\theta \in C_b(\mathbb{R}^p)$. *Then*

$$\lim_{N\to\infty} \lim_{\delta_{sim}\to 0} \mathbb{E}(\widehat{L}(\theta, \boldsymbol{Y}_n)) = L(\theta, \boldsymbol{Y}_n),$$

where the expectation is build with respect to all simulated variables and δ_{sim} *is the step*

size of the Euler method in Section 2.3.

Proof. Given $\mathbf{X}^k(t_{i-1}) \sim \pi_\theta(\mathbf{X}(t_{i-1})|\mathbf{Y}_{i-1})$, $1 \leqslant k \leqslant N$, and samples $\mathbf{X}^k_{\delta_{\text{sim}}}(t_i)$ generated by the Euler method of Section 2.3 with step size δ_{sim} conditional on $\mathbf{X}^k(t_{i-1})$, using a kernel density estimator to achieve the approximation $\hat{\pi}_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1})$, we get

$$\mathbb{E}_f(\hat{\pi}_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1})) = \mathbb{E}_f\left(\frac{1}{N}\sum_{k=1}^{N} K_{h_N}(\mathbf{X}(t_i) - \mathbf{X}^k_{\delta_{\text{sim}}}(t_i))\right) \underset{\delta_{\text{sim}}\to 0}{\to}$$

$$\frac{1}{N}\sum_{k=1}^{N} \mathbb{E}_f(K_{h_N}(\mathbf{X}(t_i) - \mathbf{X}^k(t_i))) \underset{N\to\infty}{\to} f_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1}),$$

where the expectation is taken with respect to $f_\theta(\mathbf{X}(t_i)|\mathbf{Y}_{i-1})$. The first limit is true because $K_h \in C_b(\mathbb{R}^p)$ and the weak convergence of $\mathbf{X}^{kj}_{\delta_{\text{sim}}}(t_i)$, whereas the last equation holds by the asymptotic unbiasedness of kernel density estimators in continuity points of f_θ. By this it is easy to see that the approximated resampling weights (2.29) converge to the true resampling weights. Obviously this is also true for the likelihood approximation (2.28).

Now the statement follows by induction using an iterated expectations arguement as the particle filter estimator of the likelihood function is an unbiased estimator regardless of the number of particles N (see Pitt et al. (2012) for a direct proof). □

Theorem 2.14

If the conditions in Lemma 2.13 are satisfied and $\int K^2 d\lambda < \infty$*, then*

$$\widehat{L}(\theta, \boldsymbol{Y}_n) \underset{N\to\infty, \delta_{sim}\to 0}{\overset{p}{\to}} L(\theta, \boldsymbol{Y}_n).$$

Proof. Using Markov's inequality

$$\mathbb{P}(|\widehat{L}(\theta, \mathbf{Y}_n) - L(\theta, \mathbf{Y}_n)| \geqslant \epsilon) \leqslant \frac{\mathbb{E}\left(\widehat{L}(\theta, \mathbf{Y}_n) - L(\theta, \mathbf{Y}_n)\right)^2}{\epsilon^2}.$$

Hence, in view of Lemma 2.13 and by the Cauchy-Schwarz inequality it is sufficient to show

$$\mathbb{E}\left(\widehat{L}(\theta, \mathbf{Y}_n) - \mathbb{E}(\widehat{L}(\theta, \mathbf{Y}_n))\right)^2 \to 0, \ as \ \ N \to \infty, \ \delta_{\text{sim}} \to 0.$$

However, since $\forall t_i$,

$$Var\left(\frac{1}{N}\sum_{k=1}^{N} K_{h_N}(\mathbf{X}(t_i) - \mathbf{X}^k_{\delta_{\text{sim}}}(t_i))\right) \underset{\delta_{\text{sim}}\to 0}{\to} \frac{1}{Nh_N^p}\int K^2(y) f_\theta(x + yh|\mathbf{Y}_{i-1}) dy \underset{N\to\infty}{\to} 0,$$

the statement follows by an iterated expectations argument as in the proof of Lemma 2.13. □

Remark 2.15

The direct application of Lemma 2.13 and Theorem 2.14 to CTAR is generally not possible because we do not expect that the associated transition density is a continuous function. This assumption is corroborated by the fact that for the Gaussian CTAR(1) the density of the stationary distribution is discontinuous (Brockwell (2001a)). Since the discontinuities form almost surely a (Lebesgue) null set (see Lemma 2.9), there is hope that the asymptotic of likelihood is still valid. This, however, requires further investigation.

Chopin (2004) (see also Malik and Pitt (2011)) showed that for the particle filter based estimator of the likelihood function

$$\sqrt{N}(\widehat{L}(\theta, \mathbf{Y}_n) - L(\theta, \mathbf{Y}_n)) \underset{N\to\infty}{\overset{d}{\to}} \mathcal{N}(0, \sigma_{pf}^2), \tag{2.31}$$

where σ_{pf}^2 is the particle filter variance and conditions under which $\sigma_{pf}^2 < \infty$ are given. This central limit theorem gives an idea about the error of likelihood approximation which can be interesting for model comparison as we will show in Section 2.5.

Remark 2.16

In numerical application one usually uses the logarithmic likelihood, i.e. $\log(L)$*, instead of* L*. Then the estimator* $\log(\widehat{L})$ *is no longer asymptotically unbiased. Nevertheless by Chopin (2004), a result similar to* (2.31) *is still valid for* $\log(\widehat{L})$*, where a bias of magnitude* $-\frac{\sigma_{pf}^2}{2N}$ *has to be considered.*

Theorem 2.14 serves as a basis for the assumption that $\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, \widehat{L}(\theta, \mathbf{Y}_n)$ is a meaningful estimator for our process. This is also supported by the results of Gourieroux and Monfort (1996), which could show that given explicit densities

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\to} \mathcal{N}(0, I^{-1}(\theta_0)),$$

for $n, N \to \infty$ and $\sqrt{n}/N \to 0$, where $I(\theta_0)$ is the expected information matrix at the true parameter value θ_0. However, even if $\widehat{L}(\theta, \mathbf{Y}_n)$ is a consistent estimator of a smooth likelihood function and $\hat{\theta}$ is consistent, for a finite number of particles, it is hard to optimize (2.30) by usual numerical procedures, as approximation of the likelihood by particle filtering leads to a non-smooth behavior in θ. This is because resampling particles, in fact, is sampling from a discrete distribution. Thus, by each change of θ, resampling

weights will change, and so possibly some particles are exchanged. As the replaced particles in general are not alike, $\hat{\pi}_\theta(y(t_i)|\mathbf{Y}_{i-1})$ will shift excessively.

Maximizing the resulting rough surface can be extremely problematic. Note that this also is not overcome by using common random numbers. Therefore, Campillo and Rossi (2009) include the unknown parameters θ as an additional state variable. Even if an individual noise is added to that state variable, parameter estimates in extended state space models suffer from a strong dependence on recent observations. In order to obtain an estimator that takes all observations equally into account, we prefer a classic maximum likelihood approach.

Lee (2008) proposed a tree-based resampling scheme to smooth a likelihood obtained by particle filtering with resampling by inducing significant correlation among the selected particles of consecutive runs. Figure 2.1 shows the effect of Algorithm 6 of Lee (2008) for (linear) Gaussian CAR(2) in the first dimension of the parameter space. To test

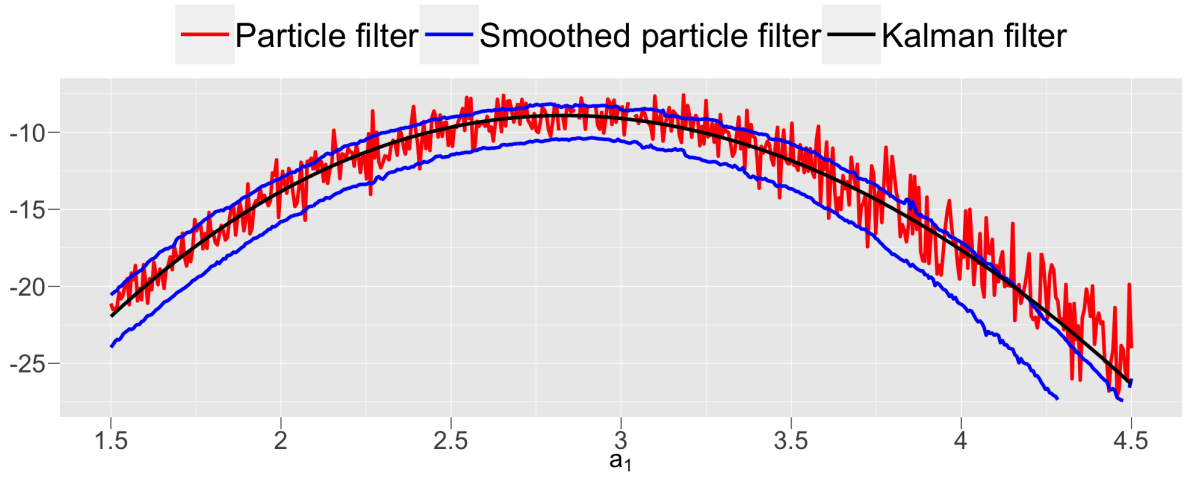

Figure 2.1: Estimated log-likelihood for a Gaussian CAR(2) model as a function of a_1 using the smoothing procedure of Lee (2008) (blue) and the vanilla particle filter (red). The true log-likelihood is shown in black. For the smoothed log-likelihood we show the estimate for two sets of common random numbers.

the general quality of the likelihood approximation for continuous-time autoregressive models, beside the smoothed estimator $\log \widehat{L}(\theta, \mathbf{Y}_n)$, we also drew the true log-likelihood $\log L(\theta, \mathbf{Y}_n)$ of the Gaussian CAR(2) calculated by the Kalman filter (cf. Section 1.5). We conclude that the smoothed particle filter results in a likelihood estimate similar to the true density of the process. This gives hope to find reasonable estimates of the parameters by numerically maximizing $\log \widehat{L}(\theta, \mathbf{Y}_n)$.

In selecting an optimization procedure to determine $\hat{\theta}$ two features of the utility function $\widehat{L}(\theta, \mathbf{Y}_n)$ should be considered. First, calculation of $\widehat{L}(\theta, \mathbf{Y}_n)$ can be computationally

intensive, which is a general drawback of many particle filter methods. The bottleneck of our method is given by the smooth resampling procedure of Lee (2008) which requires construction of a binary tree. Construction of such a tree can be done in $\mathcal{O}(NL \times \log NL)$ (see Lee (2008)). Note that (2.29) is a mixture distribution which is easy to sample from as long as K is selected appropriately. Therefore, efficient ways of optimizing in the multidimensional search space are preferred. Second, the smoothing algorithm of Lee (2008) only provides correlated likelihood estimates, but does not guarantee continuity.

A technique that can handle both is simultaneous perturbation stochastic approximation (SPSA) (see Spall (2003)[Chapter 7]). SPSA is a steepest ascent algorithm,

$$\hat{\theta}_{k+1} = \hat{\theta}_k + a_k \hat{\nabla}_{\hat{\theta}_k} L(\hat{\theta}_k, \mathbf{Y}_n),\ k = 1, \ldots, I_{\max},$$

where $\hat{\nabla}_{\hat{\theta}_k} L(\hat{\theta}_k, \mathbf{Y}_n)$ is a finite difference estimate of the gradient calculated by randomly perturbing all elements of $\hat{\theta}_k$ to obtain two measurements of $L(\cdot, \mathbf{Y}_n)$. In context of particle filtering with intractable observation density $g_\theta(y(t)|\mathbf{X}(t))$, SPSA was already used successfully by Ehrlich et al. (2015).

Table 2.1 reports the results of a simulation study for estimating the parameters of a CTAR(2) process driven by a jump diffusion, where the SPSA algorithm with the smoothed particle filter likelihood as utility function was applied. The overall results are very satisfactory. Already with 500 observations meaningful estimation results are obtained. An increase in the number of observations will also increase the precision of the estimate. Since jumps are rare events, reliable estimates for the parameters of the associated distribution can only be made with an even greater number of observations.

Table 2.1: Estimated coefficients based on 40 replicates on $[0, T]$ of a CTAR(2) with jumps. The jumps are $\pm Unif(a_\gamma, b_\gamma)$ distributed. For simulating the processes we used a bandwidth of $\delta_{sim} = 0.01$ whereas observations are taken according to the stepsize $\delta_{obs} = 1$. For estimating the likelihood we used $N = 2048$ particles.

		a_{11}	a_{12}	a_{21}	a_{22}	σ	λ	a_γ	b_γ	r_1
	true	1.5	0.5	3	1	1	0.2	0.7	2.1	0.2
T=500	mean	1.27	0.40	2.97	0.87	0.94	0.42	0.57	1.77	0.28
	bias	-0.23	-0.10	-0.03	-0.13	-0.06	0.22	-0.13	-0.33	0.08
	std. dev.	0.31	0.10	0.61	0.22	0.21	0.45	0.32	0.95	0.18
T=1500	mean	1.22	0.41	2.93	0.83	0.96	0.30	0.64	1.77	0.24
	bias	-0.28	-0.09	-0.07	-0.17	-0.04	0.10	-0.06	-0.33	0.04
	std. dev.	0.20	0.05	0.39	0.22	0.11	0.18	0.38	0.56	0.10

2.5 Application to the spot market

In Section 1.4 we discussed the characteristics of electricity prices in liberalized markets in detail. These consists of a pronounced seasonal pattern, mean reversion, and a high volatility including jumps. The temporary persistence of deviations from the average level is usually modeled by autoregressive processes. To incorporate spikes, jump-diffusions are routinely used as driving noise. However, as spikes are very short-lived they have a mean reversion rate much higher than observed for regular price levels. In estimation of a standard autoregressive process as considered in Section 1.5, this leads to an erroneous specification of the mean reversion parameters.

The common approach to overcome this is using multi-factor models (see e.g. Benth et al. (2008)). These have the drawback that separating the effect of the different components can be difficult. Instead of combining different processes Borovkova and Permana (2006) allowed the mean-reversion parameter to be a continuous function depending on the value the price. While this approach seems intuitively appealing, there is no direct approach to transfer this to a multivariate model including higher autoregressive orders. As CTAR allows for a linear interpolation of arbitrary accuracy of the mean-reversion function, the CTAR with jumps can be seen as a multivariate extension to the model of Borovkova and Permana (2006).

In Figure 2.2 the physical electricity index (Phelix) of the years 2014 to 2015 is shown. The Phelix is a stock market index representing the daily average of the day-ahead auc-

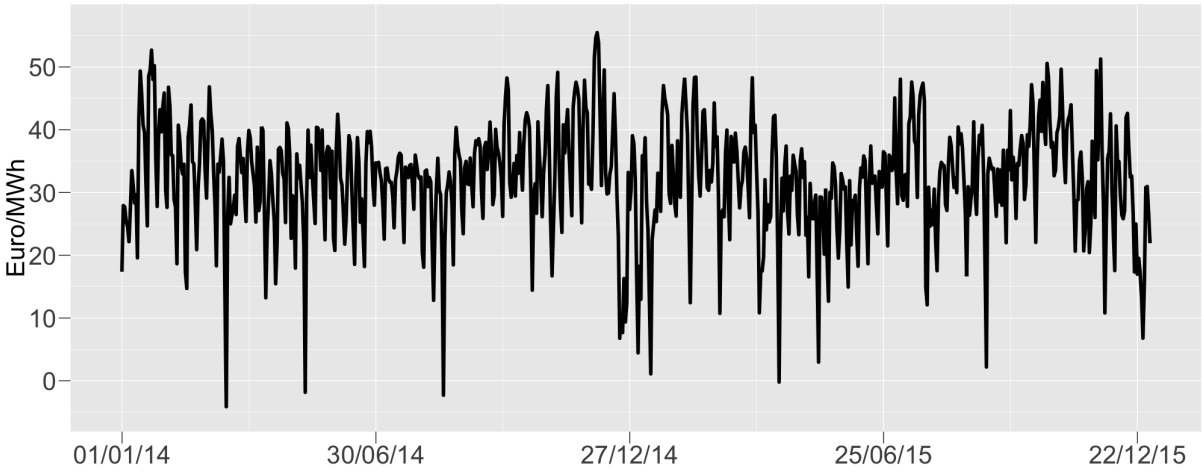

Figure 2.2: Phelix from 2014 trough 2015.

tion results (see Section 1.3 for more details). As the most important underlying on the derivatives market, accurate modeling is of great importance.

Table 2.2: Estimated parameters of the seasonality function $\Lambda(\cdot)$.

m_0	a_1	b_1	a_2	b_2	a_3	b_3
32.23	5.04	3.21	-1.58	1.22	-2.68	-2.30

The CTAR with jumps is applied to the deseasonalized time-series, i.e.

$$Y(t) = S(t) - \Lambda(t),\ t \geqslant 0,$$

where $S(t)$ is the price process and $\Lambda(t)$ is a deterministic seasonality function. Motivated by the seasonality function used in Benth et al. (2014), we assume Λ is a sum of trigonometric functions

$$\Lambda(t) = m_0 + \sum_{k=1}^{q} a_k \cos\left(\frac{2\pi t}{s_k}\right) + b_k \sin\left(\frac{2\pi t}{s_k}\right),$$

where we take $q = 3$ and the periods s_k are the dominant periods found by spectral analysis. The first three significant periods are $s_1 = 7$, $s_2 = 365$ and $s_3 = 3.5$. To estimate the parameters we used least-squares.

After subtracting $\Lambda(t)$ from the price process, we want to test if there are really nonlinear effects in $Y(t)$, which would reinforce the use of a threshold model. Therefore we use an idea of Borovkova and Permana (2006). For the Gaussian CTAR(1) as noted in Brockwell (2001a), the stationary distribution has the density

$$\pi(x) = \frac{k}{\sigma^2} \exp\left(\frac{a(x)x^2}{\sigma^2}\right), \tag{2.32}$$

where k is a normalization constant. Thus, if we replace π by an estimate of the observations marginal density (e.g. a kernel estimator), then

$$-\log(\hat{\pi}(x)) = \frac{a(x)}{\sigma^2}x^2 + \log(\sigma^2) - \log(k).$$

By this,

$$-\log(\hat{\pi}(x))' \propto a(x)x,$$

which should be a linear function if there is no nonlinear effect, i.e. $a(x) = a_1$. As (2.32) is only valid for the Gaussian process, we first filter the spikes in the data by considering those price movements as jumps which are outside of an $\pm 2\hat{\sigma}$ interval. The appropriately scaled derivative of the negative logarithmic density estimator is shown in Figure 2.3. Figure 2.3 indicates that the mean-reversion is a nonlinear function with a high value for

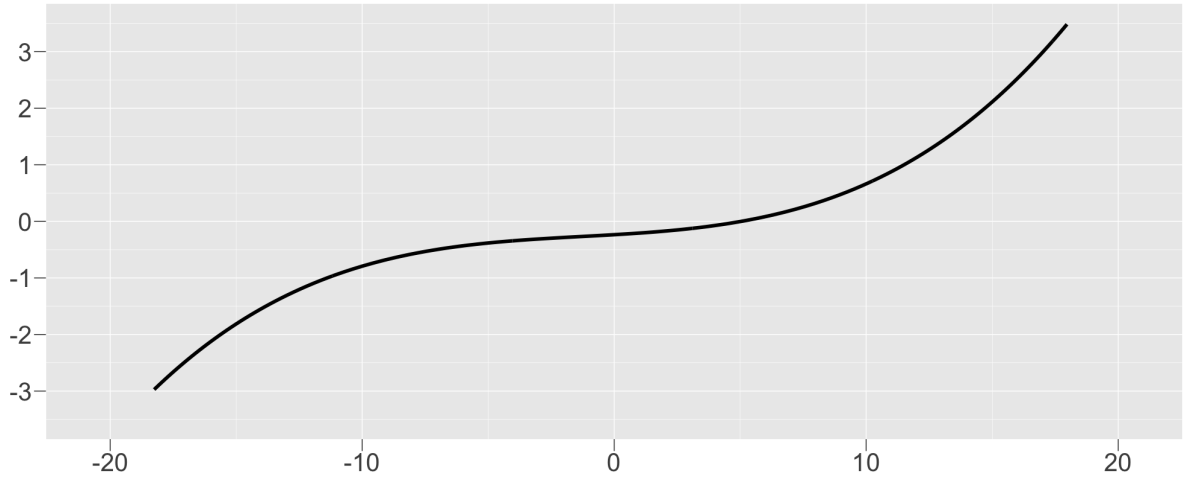

Figure 2.3: Estimated mean-reversion (drift) function.

Table 2.3: (Linear) CAR vs. CTAR for Phelix (2015)

	Phelix	CAR(2) w.j.	CTAR(2)	CTAR(2) w.j.
mean	-0.01	0.01	-0.66	-0.03
st.dev	7.08	7.21	7.97	7.86
skewness	-0.54	0.00	-0.76	-0.33
kurtosis	3.84	3.02	4.12	3.91
$\log(L)$	\	-1145.54	-1139.94	-1125.20
AIC	\	2303.08	2291.88	2268.4

extreme prices and a low value for normal price levels. This is exactly what we expect for electricity prices and indicates the use of a nonlinear model.

To test if there is really a benefit in using a nonlinear autoregressive model we finally fit a Gaussian CTAR(2), a (linear) CAR(2) driven by a jump-diffusion and a CTAR(2) with jumps to $Y(t)$, using the observations in 2015. By Figure 2.3 the use of three regimes seems reasonable. As the upper regime would include only very less observations, to get clear estimates, we restrict to one threshold instead. This is also consistent with the fact that the negative jumps are dominant for our observations. For the autoregressive order we assumed the same value as in Benth et al. (2014).

Table 2.3 compares the fitted models by means of the deviation of empirical moments, their likelihood values and the AIC. First we see that a linear CAR is not able to model the skewness in the data. This is obvious as this process is symmetric by definition as long as we use a driving process with unskewed distribution. Furthermore for the Gaussian CTAR there is a large bias in the mean. This can be explained by the dominant negative jumps in the data (see Figure 2.2), which cannot be adequately represented by

Table 2.4: Empirical standard deviation of $\log(\widehat{L})$ based on $M = 200$ simulations.

	CTAR(2)	CAR(2) w.j.	CTAR(2) w.j.
$\hat{\sigma}_{pf}$	4.29	2.70	2.61

a continuous process. In order to generate such extreme values, however, the parameters were estimated in such a way that the process is more frequently in the negative region, which leads to a downward biased mean.

For the CTAR with jumps the empirical moments suit best to our data. This is further supported by its likelihood value which outperforms the values of the other models. A more sophisticated measure to compare different models is given by the AIC. We see that the CTAR with jumps minimizes this criterion, from which we conclude that its use could be beneficial. For calculation of the likelihood in Table 2.3 we used $N = 8192$ particles and an Euler stepsize $\delta = \frac{1}{50}$. The standard deviation of the likelihood estimate was approximated numerically based on $M = 200$ simulations and can be found in Table 2.4. By these standard deviations and (2.31), the approximated log likelihood $\log(\widehat{L})$ seems to be sufficiently close to the true log likelihood $\log(\widehat{L})$. Hence, the results are well founded. Note that the bias introduced by using the logarithmic likelihood is negligible by Remark 2.16).

Chapter 3

Stochastic Modeling of Intraday PV Power Generation

This chapter is dedicated to a model for photovoltaic (PV) power generation that is able to capture intraday variation. A statistical model which makes this possible is not yet known. In Section 3.1 we explain the importance of solar power as renewable energy source and discuss the difficulties of modeling it. After an introduction to the data we consider, we discuss the need for a model that takes intraday variation of PV power generation into account. Then we define and fit such a model in Section 3.4. There we also show the benefits of the model by comparing it with an approach using only daily data. Finally we apply our model to a novel future product in Section 3.5 that allows for hedging of pure volume risk from PV power plants.

The main results of this chapter will appear in Lingohr and Müller (2019).

3.1 PV power generation

In the search for alternative energies to replace finite fossil fuels and reduce greenhouse gases, solar power plays an important role. In 2016, more additional capacity was built up by PV systems than by any other power generating technology. As a result, their share on global renewable power capacity – without hydro – has already reached 33 percent (cf. Ren21 (2017)). Since price reductions for modules have turned PV systems to be cost-competitive in the electricity market, it can be assumed that this positive trend will continue.

However, the increasing contribution of solar power to electricity generation also has disadvantages. Fluctuations in infeed due to weather conditions complicate security of

energy supplies and have significant impact on price volatility (Rintamäki et al. (2017)). In order to correctly reflect the effect of fluctuating solar infeed, we need an accurate model. Since the amount of solar power generated depends essentially on solar radiation and cloud cover, a natural approach is to use information about atmospheric conditions for modeling. This is done in the *Solis clear sky model* (cf. Mueller et al. (2004)), where radiative transfer calculations are used for forecasting. Beside the fact that this approach is complex and time consuming especially for larger geographical scale, it is not suited for financial application as it only makes statements about the short-term behavior of solar power.

Approaches to statistical modeling of solar power infeed can be found in Wagner (2012), Veraart and Zdanowicz (2016) and Benth and Ibrahim (2017), which all use methods of time series analysis. As the time series of solar infeed shows strong seasonality, the data has to be transformed first. However, this is difficult using the standard procedures, since solar power stays zero at night. This is probably the reason why the aforementioned papers are limited to daily data. Obviously, this leads to a lack in modeling.

3.2 Data description

In the following we consider the total solar power generation in Germany from Jan 1, 2015 to Dec 31, 2017. Data on the amount of electricity in gigawatt (GW) generated by solar power plants is published by the European Energy Exchange (EEX) for every hour of the day with a resolution of 15 minutes.

Figure 1.5 shows the complete time series of solar power generation for our observation period. A slight increase in the total amount of electricity generated can be confirmed over the different years. This can be attributed to the increase of installed PV capacity as already seen for the years 1990 to 2015 in Subsection 1.3.3. Figure 3.1 shows that capacity has also grow for the years 2015 to 2017, taking into account the installation of a new power plant on a daily basis. This data can also be obtained via the EEX.

Both Benth and Ibrahim (2017) and Veraart and Zdanowicz (2016) used a linear polynomial to filter the positive trend. In contrast, Wagner (2012) interpolates the final value of installed capacity of each year and used the resulting function as a normalization constant. This approach leads to a *solar efficiency process*. We do the same, but use the daily capacity levels as quotients. This removes the trend inherent in the data with maximum precision.

In Figure 1.5 we also see a strong seasonality of PV power generation on an annual basis

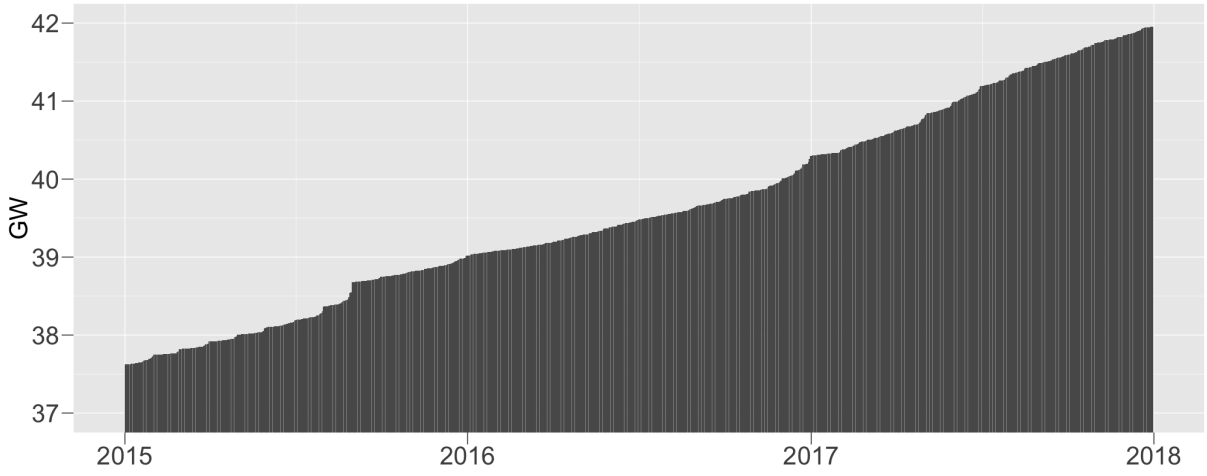

Figure 3.1: Increase in installed capacity of PV systems for 2015 trough 2017 starting with a total installed capacity at Jan 1, 2015 of 37.6 GW.

with high infeed in summer and lower infeed in winter. In addition, the yearly fluctuation is covered by a strong intraday pattern (cf. Figure 1.6). Under optimal conditions, production is maximum at noon, while it decreases towards dusk and dawn and finally falls to zero at night. Note that as solar power drops to zero overnight, it is not possible to use standard methods such as differencing or truncated Fourier series to deal with the seasonality and achieve stationarity of the process.

3.3 Intraday variation

All previous work on modeling solar power generation was limited to daily data. Of course, as we see in Figure 3.2, there is also remarkable intraday variation of the solar power infeed.

After finding a suitable model for the daily maximum infeed, Veraart and Zdanowicz (2016) recommend approximating the intraday pattern using a triangle, where the height of the top of the triangle being determined by the daily maximum and its position exactly between the times of sunrise and sunset. The daily pattern transformation of Wagner (2012) uses mean values of the ratio between the actual infeed and the daily maximum infeed for each hour of day and each month to describe the intraday pattern (cf. Subsection 1.4.3).

In Figure 3.3 we see that both methods represent only a rough approximation of the actually possible intraday pattern. Figure 3.4 also shows that although a graphical analysis of the intraday pattern could lead to the conclusion that it only varies in height (see Figure 3.2), this is not really the case. Naturally, aggregated data for a large region

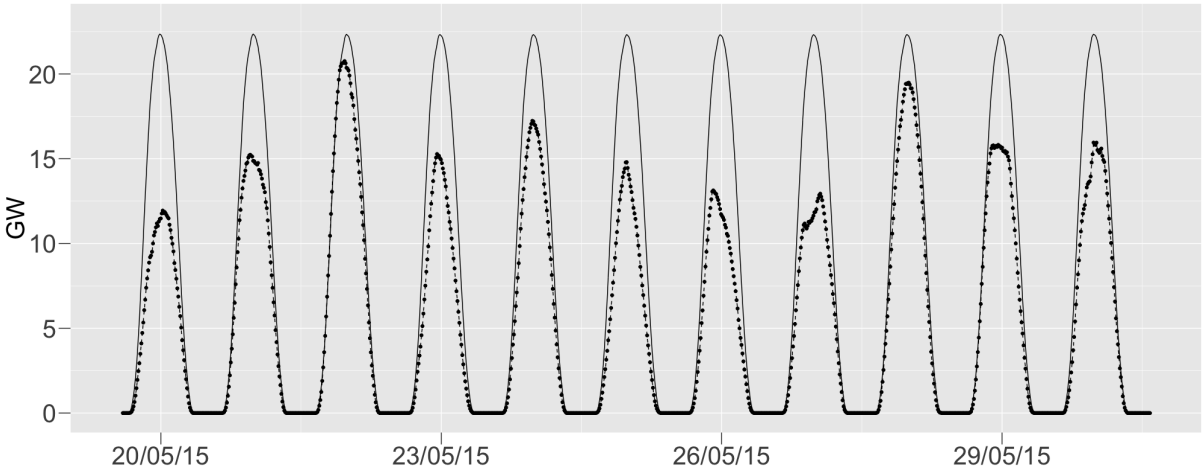

Figure 3.2: Intraday pattern of solar power generation (dotted line) and clear sky infeed (solid line) for a selected period.

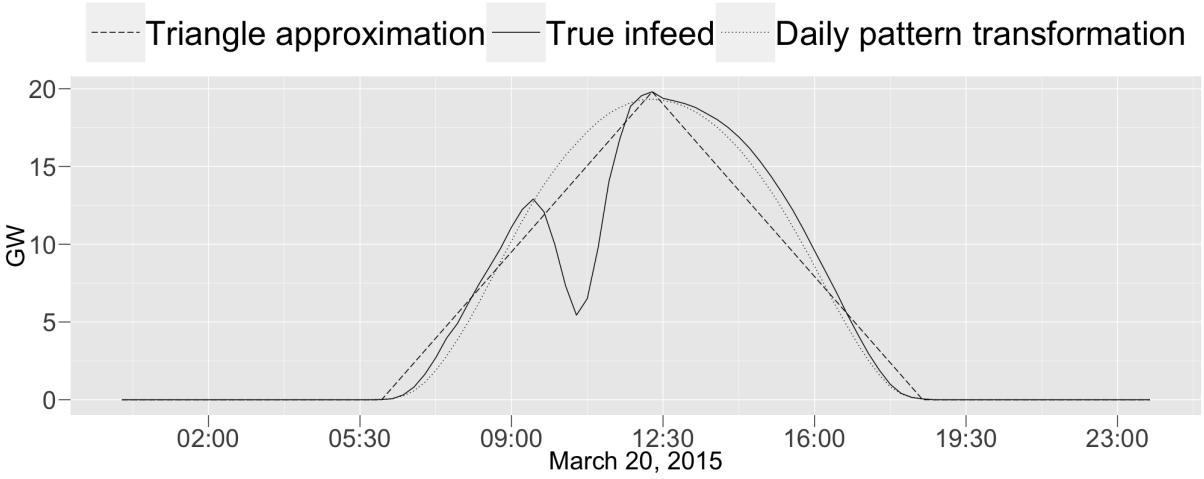

Figure 3.3: Scaled intraday patterns in comparison with true infeed on March 20, 2015.

show less extreme fluctuations at individual observation times, but if weather conditions change over the day, the daily maximum infeed can deviate from the actual point at which the sun reaches its maximum. This leads to skewed infeed curve. Accordingly, the intraday curves of Wagner (2012) and Veraart and Zdanowicz (2016) calculated on the basis of the daily maximum show a significant error especially before and after noon. For this reason we present a novel approach for solar infeed modeling in the next section, taking intraday events into account.

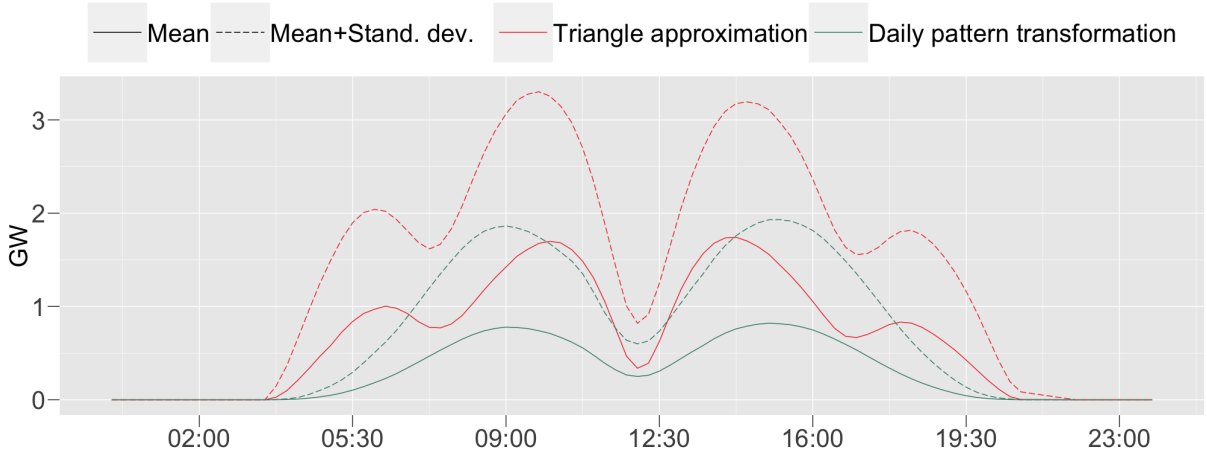

Figure 3.4: Mean and mean plus standard deviation of the absolute error between scaled intraday patterns and the actual infeed for each hour.

3.4 Clear sky model

As electricity generated from solar energy has to stay zero at night, a multiplicative relationship between the deterministic seasonality and any random variation is indicated. Obviously the same relationship holds for the installed capacity and the aforementioned process components. Therefore, we use the following representation of solar infeed, where time t is measured in days.

Definition 3.1
We describe solar power infeed by a continuous-time process $\{S(t), t \geqslant 0\}$ *with*

$$S(t) = IC(t) \times \Lambda(t) \times I(t),$$

where

$IC(t) \geqslant S(t) \geqslant 0$ *denotes installed capacity,*

$\Lambda(t) \geqslant 0$ *denotes seasonal variations, and*

$I(t) \geqslant 0$ *denotes any irregular influence.*

The detrended process

$$E(t) = S(t)/IC(t), \ t > 0, \tag{3.1}$$

is called *efficiency* (also *load factor* or *utilization rate*), where $E(t) \in [0, 1]$ must hold. The procedure described in Section 3.2 can be used to determine $IC(t)$ in case of a historical

data analysis. If Definition 3.1 is used to describe future infeed, as, for example, in forecasting or pricing of derivatives, information on the grid connection of power plants under construction can be utilized. These data can also be found on the transparency platform of the EEX. Alternatively, a simple linear model based on the historical development of solar power plants can be fitted.

If one is interested in a long term prediction, one should bear in mind that growth in installed capacity is mainly driven by political support. Therefore future solar infeed depends heavily on the expectations on government decisions. In an long term analysis of solar infeed different scenarios for $IC(t)$ should be taken into account. In such a case $IC(t)$ could be assumed to be a stochastic component. However, as we are not intending a long term prediction, we assume $IC(t)$ to be deterministic.

3.4.1 Clear sky infeed

In the search for a suitable model for $E(t)$, which takes into account intraday variation, it must be noted that for a large number of times, $E(t) = 0$. Therefore, it is not appropriate to convert $E(t) = \Lambda(t) \times I(t)$ into an additive model by logarithmizing the data. For this reason, we describe the multiplicative relationship directly.

As solar infeed is restricted to zero at night a classical fourier form for the seasonal component $\Lambda(t)$ can not be used (see also Section 3.2). A smooth estimate $\widehat{\Lambda}(t)$ that additionally accounts for outliers is described by Bacher et al. (2009). Bacher et al. (2009) used weighted quantile regression to normalize solar power. Using a high quantile level q, $\widehat{\Lambda}(t)$ approximates infeed efficiency in clear (non-overcast) sky. Hence solar models that use such an estimator for $\Lambda(t)$ are referred to as *clear sky models*, a term we will also use for the model given by Definition 3.1.

Clear sky models are usually used for decomposing global irradiance into a deterministic clear sky and a stochastic component describing transmissivity of clouds. As random fluctuations of solar infeed are generally caused by changing cloudiness, clear sky models enable a physical interpretation of $I(t)$ as *cloud cover component*. Bacher et al. (2009) determined the weights of the quantile regression by a two-dimensional smoothing kernel with a day of year and a time of day dimension consisting of Gaussian kernels ϕ as factors. Since our model should not be restricted to data of one year only we extend the kernel such that different years are incorporated.

Thus, solar radiation at time s is taken into account for time t with a weight according to

$$\Pi(t,s) = \frac{1}{n_y} \times \phi\left(\frac{d(t,s)}{\sigma_y}\right) \times \phi\left(\frac{m(t,s)}{\sigma_d}\right), \quad t,s > 0, \tag{3.2}$$

where

$n_y > 0$ is the number of years covering the data,

$d(t,s),\ m(t,s) \geqslant 0$ is the number of days resp. minutes between t and s, each identified with their position in a standard year resp. standard day, and

$\sigma_d,\ \sigma_y > 0$ is the bandwidth in the minute of day resp. day of year dimension.

The smoothing kernel (3.2) accounts for each year of the observation period with equal weight. As we do not expect a change neither in the seasonality of clear sky radiation nor in the general efficiency of solar panels (at least in the medium term) this assumption is reasonable.

Bacher et al. (2009) found that a quantile level of $q = 0.85$ gives a good match on days with clear sky all day. We follow this assessment. The choice of the bandwith σ_y and σ_d is based on visual inspection of the resulting fit. The clear sky estimator $\widehat{\Lambda}(t)$ resulting from the associated weighted quantile regression is shown in Figure 3.5 and Figure 3.2. Note that it is possible that the estimated clear sky efficiency is exceeded, i.e. $\widehat{\Lambda}(t) < E(t)$. Physically this can be explained by reflection of solar radiation and different levels of water vapour in the atmosphere (Bacher et al. (2009)). As already noted before, we do not expect a change in clear sky radiation. Therefore, $\Lambda(t)$ is assumed to be deterministic

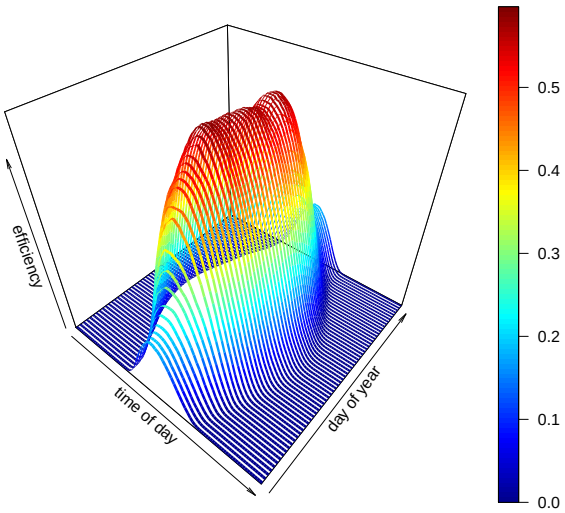

Figure 3.5: Clear sky estimator $\widehat{\Lambda}(t)$ found by weighted quantile regression with quantile level $q = 0.85$.

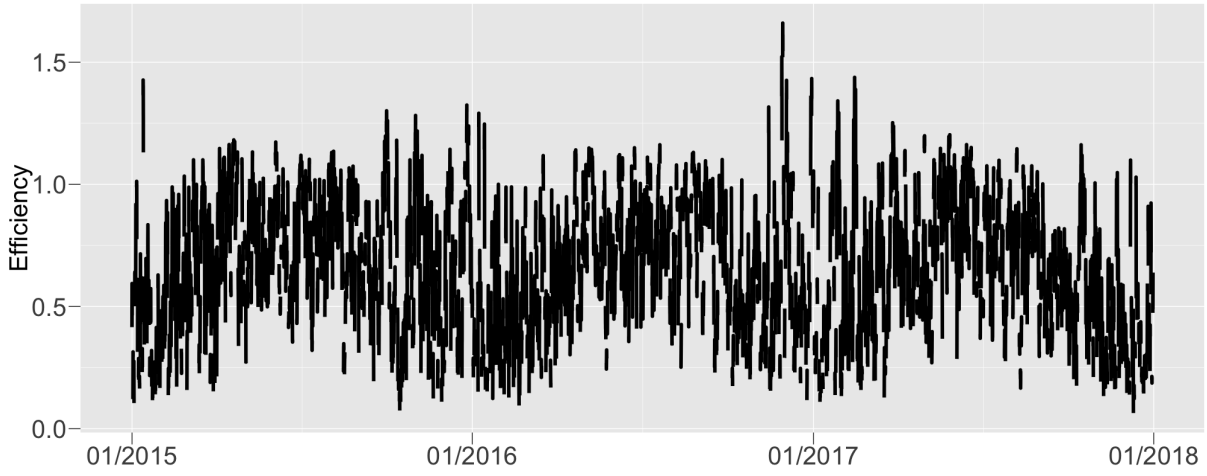

Figure 3.6: Cloud cover component $I(t)$.

and extrapolation is simply done by taking the fitted value at the same day of year and time of day.

After decomposing the solar infeed $S(t)$ into a deterministic and a stochastic part, we can now look for a suitable model of the residual component $I(t)$. This is done in the next section.

Remark 3.2
Although Benth and Ibrahim (2017) do not explicitly address the modeling of intraday data, they use the maximum sun intensity on a solar panel to display the annual seasonality of the daily maximum. The sun intensity is determined by means of physical formulas depending on place and date and can also easily be calculated for different times of day. In contrast to their physical considerations, however, a purely empirically based transformation of solar radiation is chosen to adapt the seasonality function to the data. Therefore, we see no benefit over a completely empirical approach. In addition, it is neither localized nor limited to bright hours.

3.4.2 Cloud cover component

The residual component $I(t)$ is found by dividing the solar efficiency by its clear sky estimator, i.e. $I(t) = E(t)/\widehat{\Lambda}(t)$. The process $I(t)$ therefore describes reduced infeed due to deviations from clear sky. Since such shortfalls are mainly caused by absorption of radiation by clouds, $I(t)$ is also referred to as cloud cover component.

Before we can start with the analysis of $I(t)$, it must be taken into account that for

small values of $\Lambda(t)$, $I(t)$ tends to infinity. In addition, after sunset, i.e. for $\Lambda(t)$=0, the cloud cover can no longer be determined using only infeed data. Therefore, we restrict to reliable values of $I(t)$ by removing all observations at times t, such that

$$\widehat{\Lambda}(t) < 0.35 \times \max_{u \in d_t} \widehat{\Lambda}(u) \tag{3.3}$$

where $\max_{u \in d_t} \widehat{\Lambda}(u)$ is the maximum clear sky infeed of the day associated with time t. Hence, we only take into account the times of day in which the clear sky efficiency corresponds to at least 35 % of the expected maximum daily efficiency.

In summer this is satisfied by all times between 07:00 a.m. and 5:15 p.m., while in winter only observations from 09:45 a.m. to 03:00 p.m. can be used. A similar approach is used by Bacher et al. (2009). Figure 3.6 shows the resulting process. Note that regardless of the limitation to values satisfying (3.3), by reflection of solar radiation it is still possible that the estimated clear sky efficiency is exceeded, i.e. $I(t) > 1$.

We now want to start with an analysis of the cloud cover component $I(t)$. First of all, we do not expect the general cloud cover to change within a few years. This speaks against the existence of a trend component, which can also be shown by statistical tests. In contrast, when looking at Figure 3.6, a cyclical pattern for $I(t)$ with high values in summer and low values in winter is noticeable. This is not surprising, as $\widehat{\Lambda}(t)$ has so far only taken seasonality into account due to the different positions of the sun. The additional shortfall in winter can, however, be caused by snow-covered solar panels, for example. If we examine $I(t)$ for a shorter period of time (see Figure 3.7), we also notice a parabolic evolution for most days.

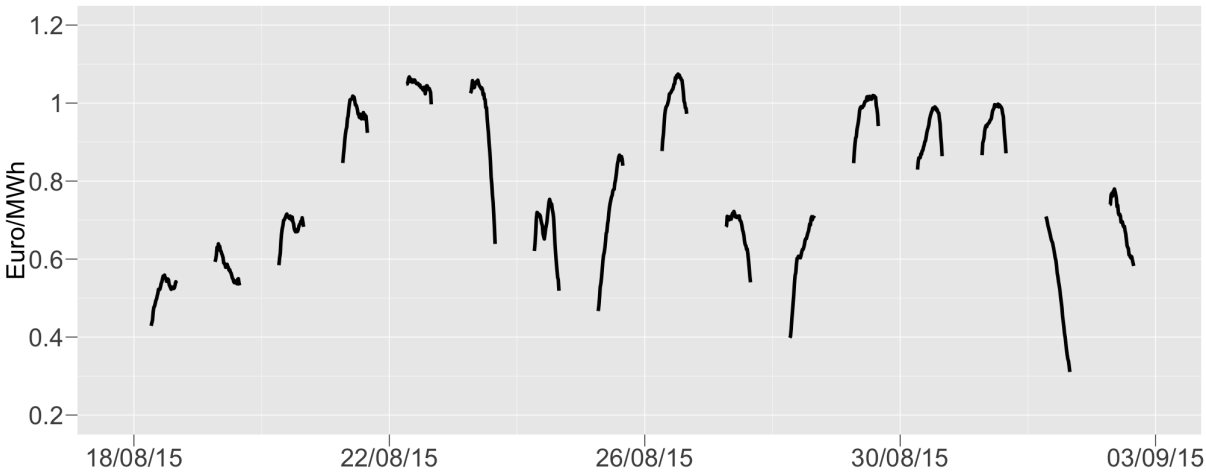

Figure 3.7: Parabolic shape of $I(t)$ on consecutive days in summer 2015.

This contradicts the assumption of a stationary process, as it can be found in many works using a clear sky model for short-term predictions (see e.g. Bacher et al. (2009) and Pedro and Coimbra (2012)). Especially in winter the daily cloud cover curves show lower values towards dawn and dusk (see Figure 3.8). By Bacher et al. (2009) non-stationarity

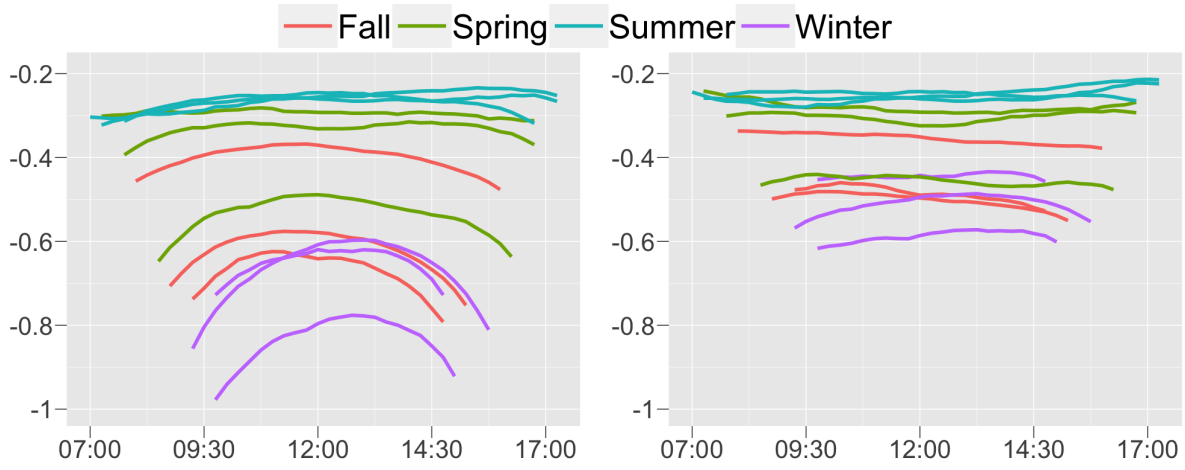

Figure 3.8: Mean value of the logged cloud cover component for each hour separated by month and colored for different seasons - before normalization (left) and after normalization (right).

during winter can be traced back to a bad estimation of $\Lambda(t)$ caused by the sparse number of clear sky observation. Comparing $\hat{\Lambda}(t)$ and the actual infeed efficiency $S(t)/\widehat{IC}(t)$ we find the estimation of the clear sky component looks quite reasonable (as far as we can say by visual inspection). Moreover even in summer the daily path of $I(t)$ does not look stationary (see Figure 3.7).

A explanation for the frequently occurring parabolic shape of $I(t)$ is given by theory of absorption. The way electromagnetic radiation is taken up by material can be described by the *Beer-Lambert law*. The Beer-Lambert law states, that for N attenuating species in a material sample of size l, the transmittance T is

$$T = \exp\left(-\sum_{i=1}^{N} \int_0^l \mu_i(z) d\right),$$

where $\mu_i(z)$, $i = 1, \ldots, N$, are attenuation coefficients. In case of uniform attenuation, this relations simplifies to

$$T = \exp(-l \sum_{i=1}^{N} \mu_i), \tag{3.4}$$

where $\tau := \sum_{i=1}^{N} \mu_i$ is called *optical depth* of the material sample.

Beer-Lambert law can be used to describe attenuation of solar radiation as it travels trough the atmosphere (Mueller et al. (2004)). In this case the optical depth depends on the *object's solar angle* (the angle between the observation object and the sun). For a slant path it is

$$\tau'(t) = \sec(\theta^{loc}(t))\tau, \tag{3.5}$$

where $\theta^{loc}(t)$ is the object's solar angle for time t and a specific location loc. As in a clear sky setup, $I(t)$ describes transmissivity of clouds, a similar form of the Beer-Lambert law has to be in effect. As $\theta^{loc}(t)$ has its maximum at solar noon and falls towards dusk and dawn, Equation (3.5) explains the parabolic shape of $I(t)$. Note that, as $\theta^{loc}(t)$ is deterministic, explaining the parabolic shape in a stochastic way leads to misspecification.

The Beer-Lambert law is stated for a pure deterministic setup. To apply it on $I(t)$ we have we have to allow for random variation over time, i.e. using stochastic processes $\mu_i(t)$ as attenuation coefficients. Thus we get for some $N \in \mathbb{N}$,

$$I(t) = \exp\left(-\sec(\theta^{loc}(t)) \sum_{i=1}^{N} \mu_i(t)\right), \tag{3.6}$$

where we set $l = 1$ as this is only a scaling factor. Relation (3.6) turns the search for a model for $I(t)$ into a search for suitable absorption coefficients. To get an idea of these, we must first filter the effect of $\sec(\theta^{loc}(t))$. Therefore we have to determine the solar angle $\theta^{loc}(t)$.

As we are considering data for complete Germany, the solar angle is based on the corresponding geographical center (see Table 3.1). In Germany the solar angle in summer is between 60° and 65°. To achieve an optimal efficiency of a PV plant it has to be inclined to $\approx 32.5°$. We will assume that most of the solar power systems are fixed and use that tilt for calculation of θ_t^{loc}. Approximative formulas for calculation of the solar angle with location coordinates as unique input parameters can be found on the website of the *Earth System Research Laboratory, Boulder (CO).*

Table 3.1: Geographical center of Germany and assumed PV tilt.

Latitude	Longitude	Tilt
$51.22°N$	$9.39°E$	30°

After the calculation of $\theta^{loc}(t)$, the effect of the attenuation coefficients is obtained according to

$$-\sum_{i=1}^{N} \mu_i(t) = \frac{\log(I(t))}{\sec(\theta^{loc}(t))} =: \bar{I}(t). \tag{3.7}$$

Interestingly, Equation (3.7) provides a physical explanation for the use of an arithmetic model of the logged data as it is often found in statistics (see e.g. Benth and Ibrahim (2017)). In contrast to standard approaches, however, the impact of $\theta^{loc}(t)$ is explicitly taken into account here. By Figure 3.8 we see that normalization of $\log(I(t))$ by means of $\sec(\theta^{loc}(t))$ contributes substantially to the removal of the parabolic daily pattern in winter and fall.

3.4.3 Attenuation processes

In Figure 3.9 we see the combined effect of all N attenuation processes $\bar{I}(t)$. There is still a seasonal effect, which is not surprising, as we have already noticed this for $I(t)$ (see Section 3.4.2).

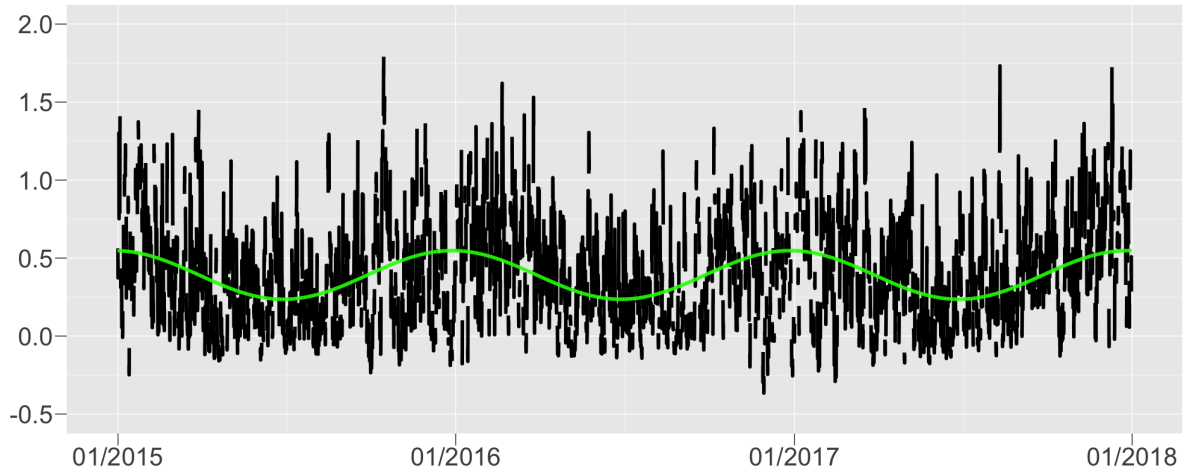

Figure 3.9: Combined attenuation processes $\bar{I}(t)$ (black) and seasonal attenuation component $\mu_1(t)$ (green).

Therefore, we suppose $\bar{I}(t)$ to be a composition of $N = 2$ components, a deterministic seasonal function

$$\mu_1(t) = a_0 + b_0 \cos\left(\frac{2\pi t}{365} + c_0\right), \tag{3.8}$$

and a stochastic component $\mu_2(t)$. This is the same model structure as used by Benth and Ibrahim (2017) to describe the logged daily maximum infeed. The parameters for $\mu_1(t)$ are found by least squares (cf. Table 3.2).

Table 3.2: Estimated parameters for the seasonal attenuation process μ_1.

a_0	b_0	c_0
0.39	0.16	0.06

The determination of a suitable model for $\mu_2(t)$ turns out to be a difficult task. First and foremost, the large number of missing observations plays a major role. Since we have no observations of the cloud cover component at night and do not receive reliable values in the morning or evening either, approximately 65 % of the observations are missing.

Continuous-time models are particularly suitable for the description of such a data set. In these, recorded data are seen as observations at possibly irregularly spaced times of an underlying stochastic process. In contrast to traditional time series, here we do not need any extra theory to handle missing data. In addition, the use of continuous-time processes is very common in finance.

To get a clue which type of stochastic process could be appropriate for our data we plot the sample autocorrelation function (ACF) of the deseasonalized $\bar{I}(t)$, i.e. $\mu_2(t) = \bar{I}(t) - \hat{\mu}_1(t)$, in Figure 3.10. Since the minimum length of consecutive observations is 6 hours, we limit ourselves to the calculation for the first 24 lags as we have one observation for each 15 min. Note that it is permissible to ignore missing data in this calculations.

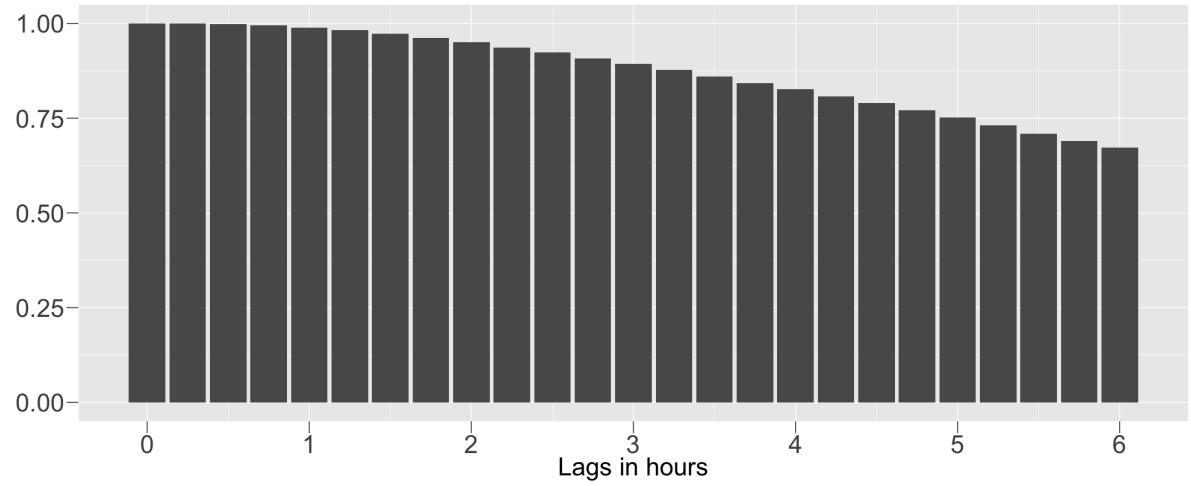

Figure 3.10: ACF of μ_2. The lags are measured in hours. The PACF has a value of 0.999 at the first lag.

We see that there is a strong dependence between the observations. The sample partial autocorrelation function (PACF) has a value of 0.999 at the first lag. This indicates the possible existence of a unit root. On the other hand, the path of $\bar{I}(t)$ (see Figure 3.9) does not correspond to what we would expect if $\mu_2(t)$ is a standard integrated process. The process looks mean reverting and in particular seems to be bounded downwards by a threshold near 0. In the positive range, the process seems less limited.

This characteristic of $\bar{I}(t)$ resp. $\mu_2(t)$ is comprehensible, since we do not assume that the clouds will change abruptly, but at the same time infeed efficiency is not unlimited

(note that a low attenuation value leads to a high infeed by (3.6)). In fact, the clear sky efficiency $\Lambda(t)$ is rarely and only slightly exceeded. Accordingly, $I(t) \not\gg 1$ resp. $\bar{I}(t) \not\gg 0$. Since we work with logarithmic data $\bar{I}(t)$, the condition $I(t) > 0$ is certainly fulfilled.

A reasonable model for $\mu_2(t)$ should therefore generate a process that is

$$\text{mean reverting,} \tag{C1}$$

$$\text{nearly integrated, and} \tag{C2}$$

$$\text{asymmetric with an lower bound.} \tag{C3}$$

Before starting to look for a suitable stochastic process, we want to draw attention to another characteristic of $\mu_2(t)$. In the study of weather-related data

$$\text{seasonal volatility} \tag{C4}$$

was often observed (see Benth and Saltyte Benth (2011) for temperature and wind speed data and Benth and Ibrahim (2017) resp. Veraart and Zdanowicz (2016) for daily maximum solar infeed data).

To test whether such behavior also exists for intraday observations and under transformation (3.7), we calculate an estimator for the variation $\sigma(t)$ of the data by forming the sums of the squared differences for each day, i.e.

$$\hat{\sigma}^2(t) = \sum_{i \in D_t} \left(\bar{I}(t_{i+1}) - \bar{I}(t_i) \right)^2, \tag{3.9}$$

where D_t contains the indices of the observations of the day belonging to time t. The estimator $\hat{\sigma}^2(t)$ given by (3.9) is the quadratic variation estimator as used in Barndorff-Nielsen and Shepard (2001).

Figure 3.11 shows clear signs of seasonal behavior with lower variation in summer than in winter. This is the same structure as mentioned in Benth and Ibrahim (2017). We follow their approach and suggest to use a deterministic truncated Fourier series to describe this cyclical dynamics,

$$\sigma^2(t) = a_1 + b_1 \cos\left(\frac{2\pi t}{365} + c_1 \right). \tag{3.10}$$

Since especially in winter $\bar{I}(t)$ can still show some unreasonable behavior towards dusk and dawn (see also Figure 3.8), we restrict D_t in (3.9) to the indices ranging from 09:15 a.m. to 03:03 p.m..

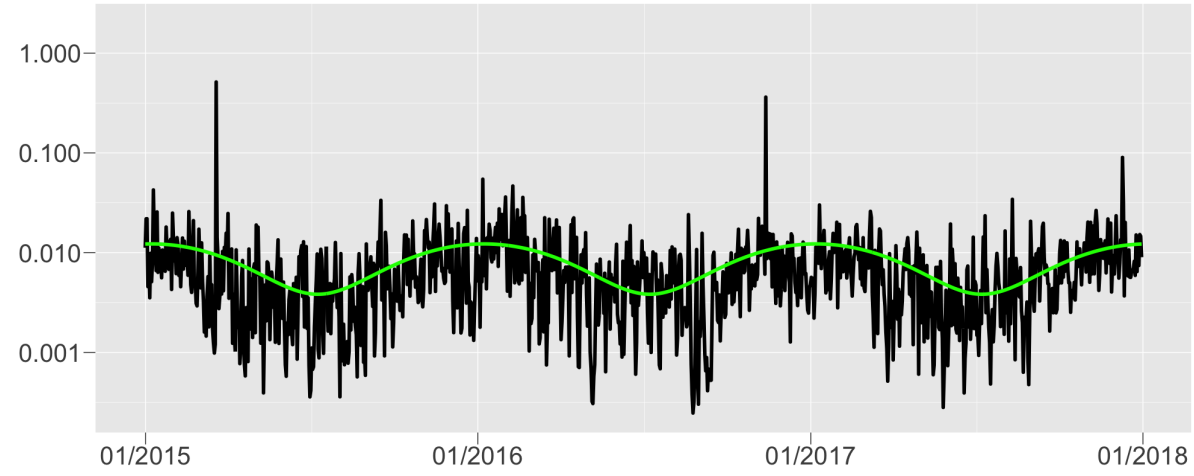

Figure 3.11: Quadratic variation estimator $\hat{\sigma}^2(t)$ (black) and fitted Fourier series (green).

Table 3.3: Least squares estimates for the seasonal variance $\sigma^2(t)$.

a_1	b_1	c_1
8e-3	4.2e-3	-1.4e-1

After summarizing the essential properties of $\mu_2(t)$ in (C1)-(C4), we now want to present a suitable stochastic process. First we remove the seasonal variation of $\mu_2(t)$ directly using

$$\tilde{\mu}_2(t) = \frac{\mu_2(t)}{\sigma(t)}, \tag{3.11}$$

where $\sigma(t)$ is as in (3.10). The corresponding parameters where found by least squares (see Table 3.3).

A class of stochastic processes with corresponding remaining characteristics is given by continuous-time threshold autoregressions (CTAR). CTAR have been introduced by Brockwell and Hyndman (1992) as a nonlinear extension to the class of continuous-time autoregressive moving average (CARMA) processes. An extension which allows for jumps was investigated in Chapter 2. However, since we assumed that $\mu_2(t)$ is a nearly integrated process, we consider here the case of the original Gaussian CTAR, i.e. the process given by Definition 2.1 with $J \equiv 0$.

A (Gaussian) CTAR(2) process $\{X(t)\}_{t\geqslant 0}$ with single threshold variable $r_1 \in \mathbb{R}$ satisfies

$$X(t) = \int_0^t X_1(t)dt, \tag{3.12}$$

where the process $\{X_1(t)\}_{t\geqslant 0}$ is the solution to the stochastic differential equation (SDE)

$$\begin{aligned} dX_1(t) = &\left[-a_{21}X(t) - a_{11}X_1(t)\right]\mathbf{1}_{\{X(t)<r_1\}} + \\ &\left[-a_{22}X(t) - a_{12}X_1(t)\right]\mathbf{1}_{\{X(t)\geqslant r_1\}} + \sigma dW(t), \end{aligned} \tag{3.13}$$

with $a_{ij} \in \mathbb{R}$, $i,j \in \{1,2\}$. Formally, a CTAR(2) process is defined by the solution $(X, X_1)(t)$ of a 2-dim. SDE (cf. Definition 2.1).

Therefore, $X(t)$ is a kind of integrated process. In addition, by separting into two regimes, an asymmetric behavior of $X(t)$ is generated. Furthermore, Stramer et al. (1996) proved that $X(t)$ is transient as long as all eigenvalues of $A(x)$ have negative real parts for each x. Therefore characteristics (C1)-(C3) are met. Obviously, a CTAR has constant variance. This is the reason why we have removed the seasonal variance in advance using (3.11). Hence, $X(t)$ is considered as a model for $\tilde{\mu}_2(t)$. To take property (C4) into account, one could also replace σ in (2.1) by $\sigma(t)$ as defined in (3.10). However, $\sigma(t)$ is then unobserved and cannot be estimated directly by (3.9).

Fitting of CTAR can be done with help of particle filtering as described Chapter 2. Here, the analytically not accessible transition densities of CTAR are approximated by kernel density estimation. Since estimating the CTAR parameters can be computational intensive, we must limit ourselves in this estimation to observations from Jan 01, 2017 to Dec 31, 2017, which still includes over 12000 data points. Estimates based on earlier years led to similar parameters. We therefore use the estimated parameters of a CTAR(2) with single threshold listed in Table 3.4 for the complete data set.

If we look at the estimation results in Table 3.4, we first notice the low values for a_{21} and a_{22}. If $a_{2i} = 0$, then $X_1(t)$ is independent of $X(t)$ for the i-th regime. According to (3.12), $X(t)$ is then actually a simple integrated process and thus (C2) is met. Despite the low estimates, an order $p > 1$ appears necessary. Besides the fact that we do not have a formal method for checking insignificance for CTAR parameters, according to Stramer et al. (1996), $X(t)$ is recurrent if $\forall i,j\ a_{ij} > 0$ and thus (C1) is met. Furthermore, $X(t)$ does not take excessive low values as described in (C3). This is because of the high mean-reversion parameter $\hat{a}_{11}$, $X_1(t)$ tends to zero fast if $X(t)$ is in the lower regime and so further negative growth of $X(t)$ is stopped.

A comparison of moments (see Table 3.5) indicates that a CTAR(2) seems to be a good choice to describe essential properties of $\tilde{\mu}_2(t)$. Only the standard deviation seems slightly too low. This is due to some positive peaks of $\tilde{\mu}_2(t)$ (see Figure 3.9) that are not adequately reproduced by the CTAR(2) process. However, these peaks do not play a significant role for the actual cloud cover component $I(t)$ after using the retransformation

Table 3.4: Estimated parameters of a CTAR(2) with seasonal volatility and single threshold.

a_{11}	a_{21}	a_{12}	a_{22}	σ	r_1
1.95	0.009	1.04	0.007	0.39	-1.78

(3.4). Note that all eigenvalues of $A(x)$ for the estimated parameters have negative real part. Therefore, by Stramer et al. (1996), stationary moments of the related CTAR(2) can be found by simulating a long time series and using ergodicity.

Table 3.5: Empirical moments of $\tilde{\mu}_2(t)$ vs. moments of fitted CTAR(2).

	mean	st.dev	skewness	kurtosis
$\tilde{\mu}_2(t)$	0.11	3.66	0.57	3.08
CTAR(2)	0.19	2.87	0.42	2.85

3.4.4 Model testing

Figure 3.12 shows a simulation of the fitted cloud cover component $I(t)$ for two different resolutions. First we see that changes in cloud cover take place slowly over a period of a few days. This is in line with our expectations for the change in weather conditions (see also Figure 3.7). Furthermore, a comparison with Figure 3.6 indicates that essential properties of $I(t)$ are meet. By defining a meaningful statistical model for $I(t)$ we are now able for the first time to reproduce not only variation of the solar power infeed between days but also within a day. For this we transform the simulated cloud cover component $I(t)$ back to solar power infeed $S(t)$ according to Definition 3.1. For the installed capacity $IC(t)$, we are using a value of 41.95 GW, which corresponds to the level as of December 31, 2017. If the model is to be used for long-term predictions, of course, this value should be replaced by a reasonable prediction of the of solar capacity expansion. Figure 3.13 illustrates that the seasonality of solar power infeed at both a yearly and a daily level is well mapped by our model. This confirms the quality of the estimator $\widehat{\Lambda}(t)$. In addition, $I(t)$ generates fluctuations in the amount of electricity fed into the grid that corresponds to the observations. In particular, deviations from the clear sky infeed are not only realized by simple scaling, but also individual behavior within a day is taken into account. This leads, for example, to a shift of the day's maximum value from the time of solar noon (cf. Section 3.2).

To measure the effect of such shifts, we consider the absolute difference between the

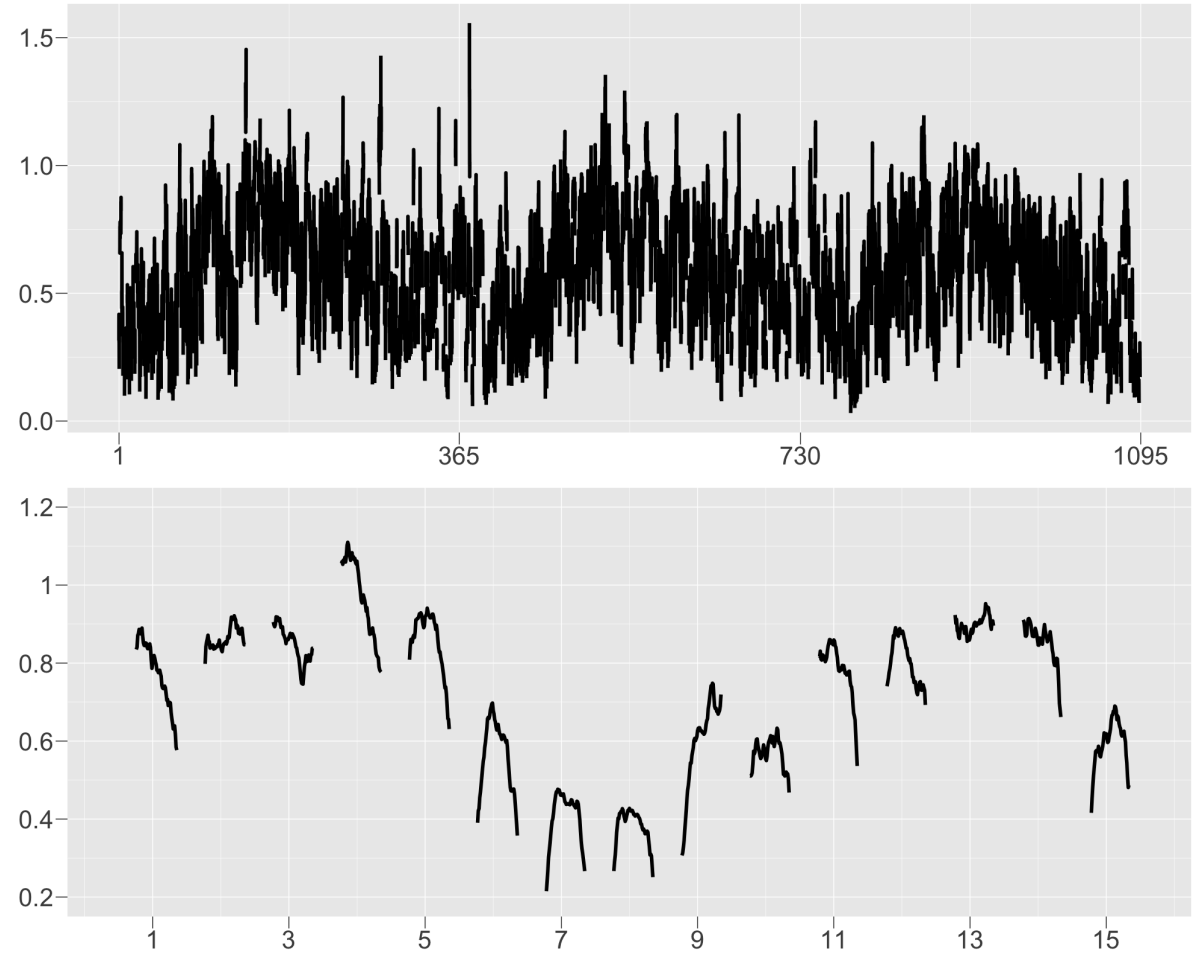

Figure 3.12: Simulated cloud cover component $I(t)$ for a period of three years (upper) and 15 consecutive days (lower). Time is stated in days.

actual infeed and the scaled clear sky infeed for each time t, i.e.

$$\left| S(t) - 41.95 \times \widehat{\Lambda}(t) \times \frac{\max\limits_{u \in d_t} S(u)}{\max\limits_{u \in d_t} \widehat{\Lambda}(u)} \right|.$$

Here, the clear sky infeed is be scaled to the same level as the true infeed, otherwise deviations due to cloudiness will cover deviations due to a shift of the daily pattern.

Figure 3.14 shows the effect of such shifts for the observed and simulated data in average for each hour. For a shifted day, the days maximum is not at solar noon. The biggest differences are therefore in the morning or in the afternoon. We see that the clear sky model realistically replicates such variations.

To emphasize the importance of defining a model for solar power infeed that respects intraday effect, we also want to draw a comparison to the model of Wagner (2012), which

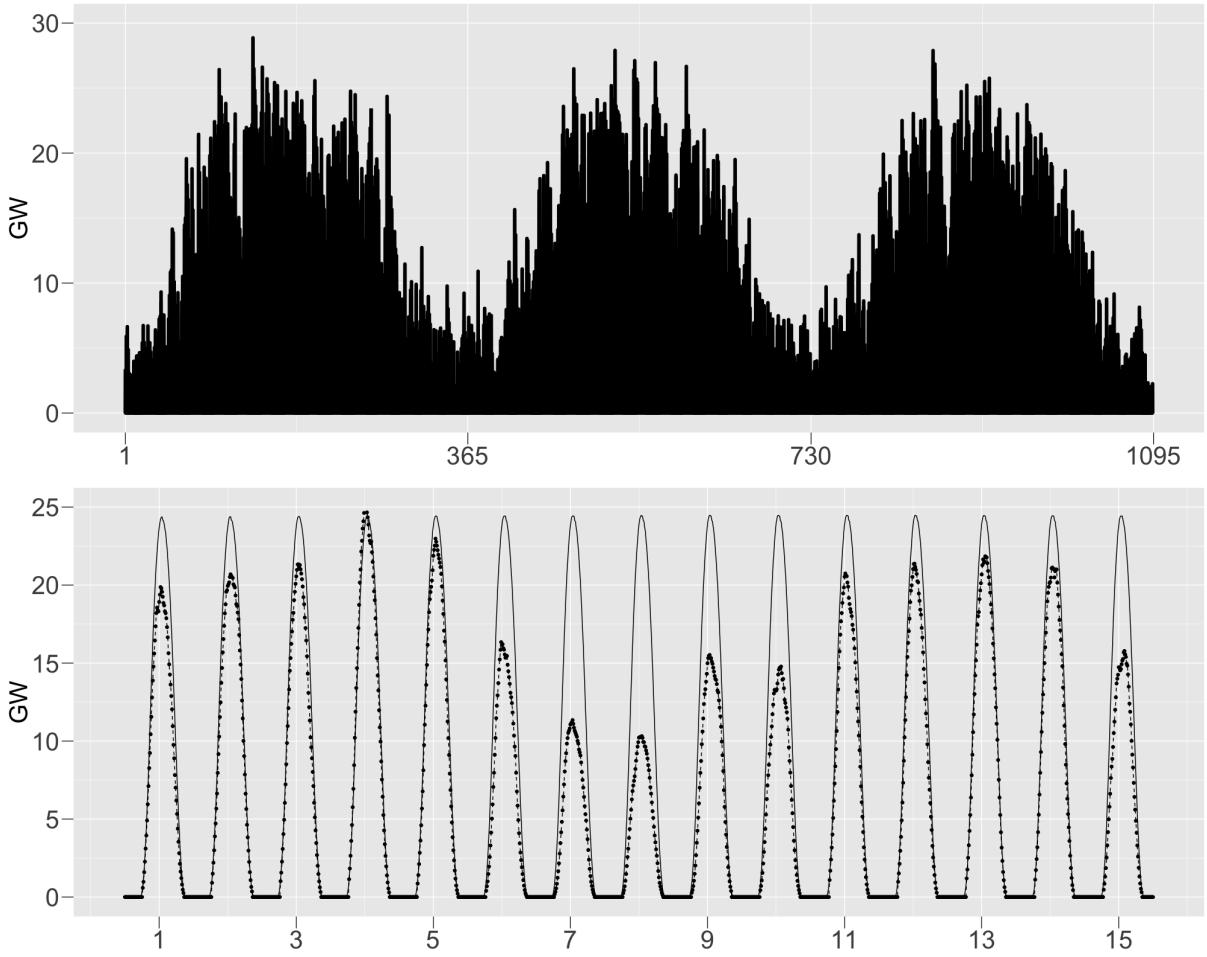

Figure 3.13: Simulated solar power infeed $S(t)$ for a period of three years (upper) and 15 consecutive days (lower, dotted line) together with clear sky infeed (lower, solid line). Time is stated in days.

is up to our knowledge the most accurate statistical model for describing solar power infeed not only on a daily scale. Therefore, we fitted the daily maximum process as defined in Wagner (2012) and used the daily pattern transformation (see Section 3.2) to get intraday infeed curves. Since the main motivation for defining a solar infeed model that allows for individual daily structures laid in the observed deviations of the infeed from the regular pattern, we also calculated the effect of a shift for the model of Wagner (2012) (see Figure 3.14).

As to be expected, such effects are insufficiently reflected by this model. Interestingly, the simulation based on Wagner (2012) takes some values which result in unreasonably high infeed (see Table 3.6). This could be due to the fact that the time series of daily maxima is described by a symmetrical process. In our approach, skewness in the data is taken into account by using a threshold process.

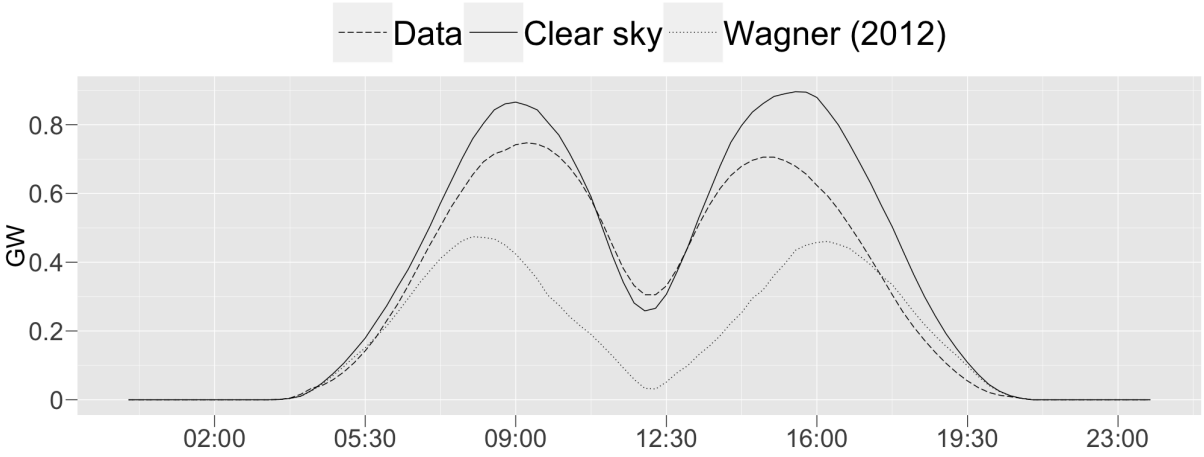

Figure 3.14: Deviations of the daily infeed patterns from the scaled clear sky infeed patterns.

Table 3.6: Empirical upper tail quantile of $E(t)$ restricted to daily values at 12 p.m..

	99%	99.9%	99.95%
Data	0.658	0.677	0.679
Clear Sky	0.616	0.701	0.736
Wagner (2012)	0.699	0.780	0.794

An excessively highly valued solar power infeed can be problematic, especially as a component in an electricity price model. This is because high solar power infeed can lead to extreme prices (cf. Chapter 4).

3.5 PV power futures

Increased infeed from renewable energy sources necessitates the introduction of new products for electricity markets. Since 2015 the EEX lists the class of *Energiewende products*, which are derivatives for hedging against weather related uncertainties. As part of the Energiewende products, EEX has introduced a future on the average load factor of wind production (see Benth and Pircalabu (2018) for a analysis of wind power future prices). The wind power future provides the opportunity to hedge pure volume risk in wind power generation.

Under the Gesetz für den Ausbau erneuerbarer Energien (Renewable Energy Sources Act) (EEG), renewable energies can be subsidized by receiving a market premium additional to the exchange price (cf. Section 1.3). The operator of a subsidized power plant is therefore no longer exposed to any price risk. This is why futures for renewable energies

that only reflect volume risks are reasonable. Since solar power has the second largest installed capacity in Germany, such a product also seems necessary to hedge volume risk of PV power plants.

Thus, we define a new type of contract which we call PV power future with future price $F_{PV}(t, \tau_1, \tau_2)$ at time t and payoff

$$\frac{100}{\tau_2 - \tau_1} \sum_{u=\tau_1}^{\tau_2} E(u), \tag{3.14}$$

for a measurement period $[\tau_1, \tau_2]$, $\tau_1 \leqslant \tau_2$. The future contracts are settled in Euro at the end of the measurement period. As for wind power futures, the holder of the PV power future receives the average load of solar power stations across Germany.

In line with the practice for electricity markets (see Benth et al. (2008)), we assume that the future is valuated on the basis of rational expectations, i.e.

$$F_{PV}(t, \tau_1, \tau_2) = \mathbb{E}_{\mathbb{P}} \left(\frac{100}{\tau_2 - \tau_1} \sum_{u=\tau_1}^{\tau_2} E(u) \middle| \mathcal{F}_t \right) + RP(t, \tau_1, \tau_2), \tag{3.15}$$

were $\mathbb{E}_{\mathbb{P}}$ is the expectation with respect to the objective probability $\mathbb{P}$ and $RP(t, \tau_1, \tau_2)$ is a risk premium.

In accordance with the theory of mathematical finance, the risk premium is usually expressed in terms of a risk-neutral measure $\mathbb{Q}$, i.e. an equivalent martingale measure. Due to the successful application of stochastic differential equations in the pricing of derivatives, continuous-time processes and first and foremost the Brownian motion have become popular to model problems in finance. Therefore, a common choice of $\mathbb{Q}$ (see e.g. Benth et al. (2007)) is given via the Girsanov transform

$$\frac{d\mathbb{Q}}{d\mathbb{P}} | \mathcal{F}_t = \exp \left(\int_0^t \frac{\theta(s)}{\sigma(s)} dW(s) - \frac{1}{2} \int_0^t \frac{\theta^2(s)}{\sigma^2(s)} ds \right), \tag{3.16}$$

where $\theta(t)$ is a real-valued, bounded and piecewise continuous function on $[0, T]$ called the *market price of risk* and expressing the risk premium (see Benth et al. (2008)[1.5.2] for details).

We now want the use the proposed model of Section 3.4 to analyze the PV power future price F_{PV}. First, note that the SDE (2.1) can represented as one-dimensional SDE (Brockwell (1994)) and so the class of risk-neutral probabilities given by (3.16) simply adds a drift of size $\theta(t)dt$ to $dX_1(t)$ which is easy to handle. In a concrete application $\theta(t)$ could be calibrated to observed future prices. Since such futures are not available at

the moment, for the following studies we set $\theta \equiv 0$, i.e. $\mathbb{Q} = \mathbb{P}$.

Proposition 3.3
The future price $F_{PV}(t, \tau_1, \tau_2)$ of a PV power future is

$$F_{PV}(t, \tau_1, \tau_2) = \frac{100}{\tau_2 - \tau_1} \sum_{u=\tau_1}^{\tau_2} \Lambda(u) e^{-sec(\theta^{loc}(u))\mu_1(u)} \mathbb{E}_{\mathbb{Q}} \left(e^{-sec(\theta^{loc}(u))\sigma(t)X(u)} | (X, X_1)(t) \right).$$

Proof. The representation of $F_{PV}(t, \tau_1, \tau_2)$ follows directly from (3.15) inserting $E(t)$ as defined in Section 3.4 and the fact that $(X, X_1)(t)$ is a Markov process (see Stramer et al. (1996)). □

By Proposition 3.3, the use of the clear sky model in future pricing requires the existence of exponential moments of a CTAR(2) process. Since $X(t)$ is unbounded for both regime, we have to proof this condition separately.

Theorem 3.4
For any $t \geqslant 0$, it holds that $\mathbb{E}_{\mathbb{Q}}(e^{cX(t)}) < \infty$ for any $c \in \mathbb{R}$, where $X(t)$ is a (Gaussian) CTAR(2) process.

Proof. As we do not have an analytic expression of the solution to the SDE (3.13), to calculate the exponential moment of a CTAR process we use an idea of Brockwell (1994). Given the stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0}, \mathbb{P})$, we know that the Gaussian process

$$dX_1(t) = \sigma dW(t),$$

is a weak solution to the SDE (3.13) under the measure

$$\mathbb{Q} = \int_A M(t, X_1) d\mathbb{P}, \ \forall A \in \mathcal{F}_t,$$

where

$$M(t, X_1) = \exp\left(\int_0^t \frac{H(s, X_1)}{\sigma} dW(s) - \frac{1}{2} \int_0^t \frac{H^2(s, X_1)}{\sigma^2} ds \right), \tag{3.17}$$

with $H(t, X_1) = G(t, X_1) + \theta(t)$ and $G(t, X_1)$ is as in Theorem 2.4. Therefore

$$\mathbb{E}_{\mathbb{Q}}(e^{cX(t)}) = \mathbb{E}_{\mathbb{P}}(e^{cX(t)} M(t, X_1)).$$

Using Hölder's inequality we find

$$\mathbb{E}_{\mathbb{Q}}(e^{cX(t)}) \leqslant \left[\mathbb{E}_{\mathbb{P}}\left(e^{3cX(t)}\right)\mathbb{E}_{\mathbb{P}}\left(e^{3\int_0^t \frac{H(s,X_1)}{\sigma}dW(s)}\right)\mathbb{E}_{\mathbb{P}}\left(e^{-\frac{3}{2}\int_0^t \frac{H^2(s,X_1)}{\sigma^2}ds}\right)\right]^{\frac{1}{3}}. \tag{3.18}$$

W.l.o.g. we assume $(X, X_1)(0) = 0$ almost surely in the following. As σ is bounded and deterministic

$$X(t) = \int_0^t X_1(u)du = \int_0^t \int_0^u \sigma dW(s)dsdu = \int_0^t \sigma W(u)du,$$

is a Gaussian process under $\mathbb{P}$ with mean

$$\mathbb{E}_{\mathbb{P}}(X(t)) = 0,$$

and variance

$$Var_{\mathbb{P}}\left(X(t)\right) = \int_0^t \int_0^t \mathbb{E}_{\mathbb{P}}\left(\sigma^2 W(u_1)W(u_2)\right)du_1du_2 =$$

$$= \int_0^t \int_0^t \sigma^2 \min\left(u_1, u_2\right)du_1du_2 = \sigma^2\left(\frac{1}{6}t^3 + \frac{1}{2}t^2\right).$$

Hence, we get

$$\mathbb{E}_{\mathbb{P}}(e^{3cX(t)}) = \exp\left(3c\mathbb{E}_{\mathbb{P}}(X(t)) + \frac{9c^2}{2}Var_{\mathbb{P}}(X(t))\right) < \infty.$$

Since the third factor on the right hand side of (3.18) is trivially bounded by 1, it stays to proof

$$\mathbb{E}_{\mathbb{P}}\left(e^{3\int_0^t \frac{H(s,X_1)}{\sigma}dW(s)}\right) < \infty. \tag{3.19}$$

For this first note that

$$\mathbb{E}_{\mathbb{P}}\left(e^{3\int_0^t \frac{H(s,X_1)}{\sigma}dW(s) - \frac{1}{2}\int_0^t \frac{36H^2(s,X_1)}{2\sigma^2}ds + \frac{1}{2}\int_0^t \frac{36H^2(s,X_1)}{2\sigma^2}ds}\right) \leqslant \left[\mathbb{E}_{\mathbb{P}}\left(e^{\frac{1}{2}\int_0^t \frac{36H^2(s,X_1)}{\sigma^2}ds}\right)\right]^{\frac{1}{2}},$$

by Hölder's inequality and the fact that $e^{6\int_0^t \frac{H(s,X_1)}{\sigma(s)}dW(s) - \frac{1}{2}\int_0^t \frac{36H^2(s,X_1)}{\sigma^2(s)}ds}$ is a martingale. This is because

$$\left(\frac{6H(s,X_1)}{\sigma}\right)^2 \leqslant K_t\left(1 + \max_{s\in[0,t]} X_1^2(s)\right),\ s \leqslant t, \tag{3.20}$$

for some $K_t > 0$ as a_{ij}, $i, j \in \{1, 2\}$, σ and $\theta(t)$ are bounded. This means that $\frac{6H(s,X_1)}{\sigma}$ satisfies the linear growth condition of Lemma 2.3. By Hölder's inequality, (3.20) and Doob's martingale inequality for a sequence $0 = t_0 < t_1 < \cdots < t_n = t$ with $|t_i - t_{i-1}| = \Delta$

we conclude

$$\mathbb{E}_{\mathbb{P}}\left(e^{\frac{1}{2}\int_0^t \frac{36H^2(s,X_1)}{\sigma^2}ds}\right) \leqslant \prod_{i=1}^{n}\left[\mathbb{E}_{\mathbb{P}}\left(e^{\int_{t_{i-1}}^{t_i} \frac{36H^2(s,X_1)}{\sigma^2}ds}\right)\right]^{\frac{1}{2}} \leqslant$$

$$\prod_{i=1}^{n}\left[\mathbb{E}_{\mathbb{P}}\left(e^{K_t(1+\max_{s\in[0,t]} X_1^2(s))\Delta}\right)\right]^{\frac{1}{2}} \leqslant 4^n e^{K_t t}\prod_{i=1}^{n}\mathbb{E}_{\mathbb{P}}\left(e^{K_t\Delta X_1^2(s)}\right),$$

which is finite provided that $\Delta < \left(2K_t\sqrt{t}\sigma\right)^{-1}$ as X_1 is Gaussian under $\mathbb{P}$. □

Sample paths of a CTAR(2) process can be simulated consistently by an Euler approximation (cf. Chapter 2). We use this for computing the conditional moment in Proposition 3.3 numerically.

Figure 3.15 shows the distribution of possible payoffs of the PV power future using different month as measurement period. Each distribution is approximated based on 50,000 simulated path of $X(t)$, which are started two month before τ_1 such that the influence of the initial $(X, X_1)(0) = 0$ is negligible (cf. Figure 3.16).

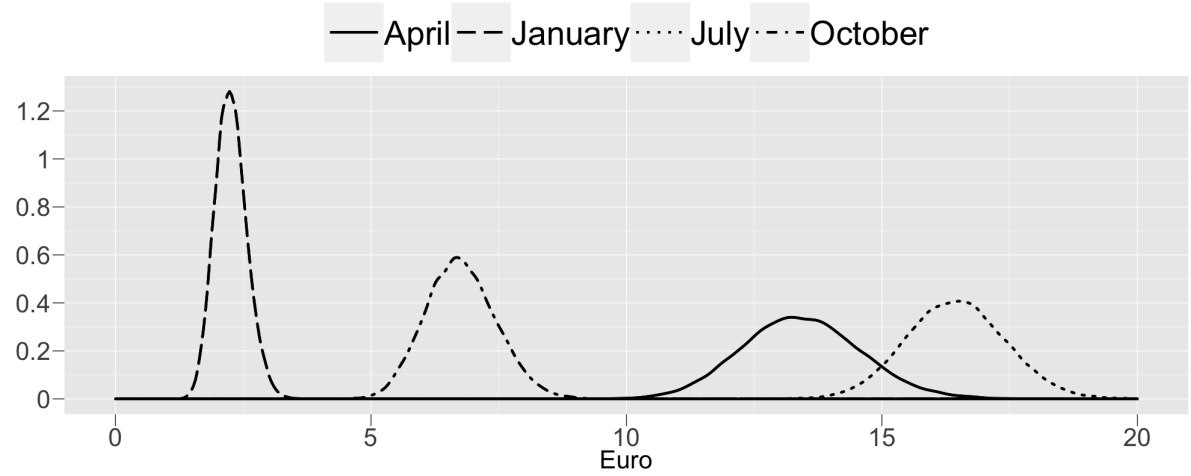

Figure 3.15: Simulated distribution of the payoff of a PV power future for four example month.

We see that in winter there is much less variation in the payoff than in summer. This indicates that hedging the volume risk for month with high clear sky infeed generally contribute more to reducing the risk exposure of a PV power plant. This is due to the fact that, in absolute terms, in summer even small variations can have a great effect. The relatively high clear sky infeed combined with larger variations lead to the highest spread in spring.

Note that by the ergodicity of $(X, X_1)(t)$ the mean value of each distribution in Figure 3.15 is equal to $F_{PV}(t, \tau_1, \tau_2)$ for $\tau_1 - t$ large enough. This implies a clear seasonality in the future price, which is a common feature in commodity markets.

Another feature of future prices in commodity markets found by Samuelson (1965) is that "the variations of distant maturity futures are lower than nearby futures prices". This known as the *Samuelson effect*. As our cloud cover component is mean-reverting we also expect that $F_{PV}(t, \tau_1, \tau_2)$ exhibits a volatility that is decreasing with time to delivery. Figure 3.16 shows that this expectation is meet. The volatility of the future price was estimated based on 1000 simulations. As expected, a shorter measurement period is

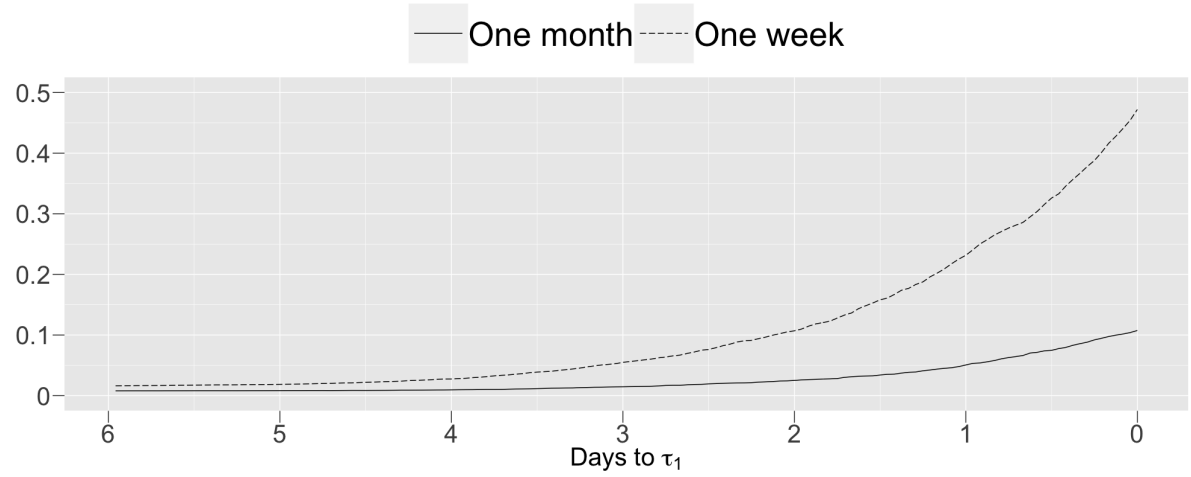

Figure 3.16: Visualization of the Samuelson effect with time measured in days prior to τ_1, July 1st for measurement periods of different length.

also associated with a higher volatility of the future. This is because extreme weather conditions do not remain long-term and therefore there is greater uncertainty regarding the average efficiency achieved for short periods of time. Using Figure 3.16 we also see that $(X, X_1)(t)$ has an influence on $F_{PV}(t, \tau_1, \tau_2)$ only a few days prior to τ_1. This is in line with the fact that the weather can only be accurately predicted for a few days in advance.

Chapter 4

CII Processes for Modeling Electricity Prices with regard to Renewable Power Generation

The impact of renewable power generation on the electricity price has caused statisticians to seek for approaches to incorporate associated information into their price models (cf. Section 1.4). In this chapter we demonstrate that this is possible by directly varying the parameters of the distribution of a process increments. This is a very flexible method and also seems an obvious approach from a statistical perspective.

In Section 4.1 we will show how such a process is constructed, prove the consistency of nonparametric moment estimates and present a test to check if there is really a dependency on external data. The quality of a parameter estimation procedure and the aforementioned test is numerically verified. Afterwards, in Section 4.2, we define an electricity price model based on the new process that incorporates information on renewable power generation explicitly. We fit the model to the German intraday market and check its quality. Finally, the model is used in Section 4.4 to analyze German Intraday Cap/Floor Futures prices.

4.1 CII processes

To make the behavior of a stochastic process dependent on an external variable, we take up an idea in Cont and Tankov (2004)[Chapter 14]. They showed how a stochastic process whose behavior should change as time evolves can be defined using the theory of additive processes (cf. Sato (1999)). Additive processes, also known as processes with independent increments, represent a useful class in application, since they preserve almost all the

tractability of Lévy processes, but allow for a flexible choice of the increment distribution depending on time.

To construct an additive process, Cont and Tankov (2004) used a triplet $(\Gamma(t), A(t), \nu(t))_{t\geqslant 0}$, where

$$\left.\begin{aligned} &\Gamma(t) := \textstyle\int_0^t \gamma(s)ds, \\ &A(t) := \textstyle\int_0^t \sigma^2(s)ds, \\ &\nu(t,B) := \textstyle\int_0^t \mu(s,B)ds, \;\; B \in \mathcal{B}(\mathbb{R}) \end{aligned}\right\}, \tag{4.1}$$

for functions $\gamma : \mathbb{R}^+ \to \mathbb{R}$, $\sigma : \mathbb{R}^+ \to \mathbb{R}$, and a family of Lévy measures $\{\mu(t)\}_{t\geqslant 0}$. Under rather weak conditions (see Sato (1999)[Theorem 9.8]), this triplet is the *generating triplet* of an infinitely divisible distribution, which represents the distribution of an additive process X. Here, the behavior of the process X is specified by the *local characteristics* $(\gamma(t), \sigma^2(t), \mu(t))_{t\geqslant 0}$ which are assumed to be deterministic.

We extend the approach of Cont and Tankov (2004) by allowing the local characteristics to be stochastic. This allows the distribution of X to depend on an external random variable. Since for additive processes the change of the distribution must done in a deterministic way, however, this leads to another class of processes for X called conditionally independent increment (CII) processes.

CII processes were introduced previously (see Jacod and Shiryaev (2002)[II.6]) to investigate convergence of random variables. In Äit Sahalia and Jacod (2014) they appear in this context for estimating the variance in a stochastic volatility model. Theoretical properties of CII processes have also been investigated much earlier (Grigelionis (1975)). However, although they promise a very flexible modeling approach, to our knowledge direct use of such processes for practical application has not yet been investigated.

4.1.1 Existence

We want to enable the behavior of X to be controlled by an external process I. Therefore, on a probability space $(\Omega^I, \mathcal{F}^I, \mathbb{P}^I)$, we assume that

- $\gamma = \{\gamma(t)\}_{t\geqslant 0}$ and $\sigma = \{\sigma(t)\}_{t\geqslant 0}$ will be real-valued stochastic processes, and
- $\mu = \{\mu(t)\}_{t\geqslant 0}$ a family of random measures on $\mathbb{R}$.

These characteristics express the information given by a real-valued stochastic process $I = \{I(t)\}_{t\geqslant 0}$, called *external information*, by being adapted to it, i.e.

- γ, σ and $\mu(B)$, $B \in \mathcal{B}(\mathbb{R})$, will be adapted to the natural filtration $\mathbb{F}^I = \{\mathcal{F}_t^I\}_{t\geqslant 0}$, where $\mathcal{F}_t^I = \sigma\{I(s) : s \leqslant t\} \subset \mathcal{F}^I = \sigma(I(s) : s \geqslant 0)$.

Such a triplet (γ, σ^2, μ) is briefly referred to as I-adapted triplet in the following.

Since for an additive process the generating triplet has to be non-stochastic, existence in the stochastic case is not covered by the previous results. However, since for each path $I(\omega)$ an additive process can be defined, X can be constructed by specifying the distribution $X|I$, i.e. the distribution of X under $\mathcal{F}^I$.

Theorem 4.1
There exists a stochastic process $X = \{X(t)\}_{\geqslant 0}$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ such that $\mathbb{P}$-a.s. $X|I(\omega) \stackrel{d}{=} Z^{(I(\omega))}$, where $Z^{(I(\omega))}$ is an additive process with local characteristics $(\gamma(t)(\omega), \sigma^2(t)(\omega), \mu(t)(\omega))_{t\geqslant 0}$, if

i) γ, σ^2 and $\mu(A)$, $A \subset \mathcal{F}$, are progressively measurable,

ii) $\int_0^t |\gamma(s)|ds < \infty$, $\int_0^t \sigma^2(s)ds < \infty$, $\mathbb{P}$-a.s. $\forall t \geqslant 0$, and

iii) $\int_0^t \int_{\mathbb{R}} (x^2 \wedge 1)\ \mu(s, dx)ds < \infty$, $\mathbb{P}$-a.s. $\forall t \geqslant 0$,

for (γ, σ^2, μ) being an I-adapted triplet. The process X is then called an I-conditionally independent increment (I-CII) process.

Proof. Let $\Omega^* = D(\mathbb{R}^+, \mathbb{R})$ be the Skorokhod space and $\mathcal{F}^*$ the canonical σ-algebra. X^* is defined as the canonical process $X^*(t)(\omega^*) = \omega^*(t)$. We consider the product space $(\Omega, \mathcal{F}) = (\Omega^I \times \Omega^*, \mathcal{F}^I \otimes \mathcal{F}^*)$ equipped with a suitable filtration $\mathbb{F}$, where all considered processes are extended in the usual way to Ω, for example $X(\omega^I \times \omega^*) = X^*(\omega^*)$.

Under the above conditions we know by Sato (1999)[Theorem 9.8], that for $\omega^I \in \Omega^I$ a probability measure $\mathbb{P}^*_{\omega^I}(d\omega^*) = \mathbb{Q}(\omega^I, d\omega^*)$ on $(\Omega^*, \mathcal{F}^*)$ exists (a.s.), such that X^* is additive with local characteristics $(\gamma(t)(\omega), \sigma^2(t)(\omega), \mu(t)(\omega))_{t\geqslant 0}$.

Now, if $\mathbb{Q} : \Omega^I \times \mathcal{F}^* \to [0, 1]$ is a stochastic kernel, then

$$\mathbb{P}(d\omega^I, d\omega^*) = \mathbb{P}^I(d\omega^I)\mathbb{Q}(\omega^I, d\omega^*),$$

is a probability measure on $(\Omega, \mathcal{F})$ (Klenke (2008)[Korollar 14.23]). Obviously, $\mathbb{P}(X \in B|I(\omega^I \times \omega^*)) = \mathbb{P}^*_{\omega^I}(B^*)$, $B = B^I \times B^* \subset \mathcal{F}$, then holds, i.e. X is a I-CII process.

It remains to prove that $\mathbb{Q}$ is a stochastic kernel. First, $\forall \omega^I \in \Omega^I$ $B^* \mapsto \mathbb{Q}(\omega^I, B^*)$, $B^* \in \mathcal{F}^*$, is a probability measure on $(\Omega^*, \mathcal{F}^*)$ by construction. To show that $\omega^I \mapsto$

$\mathbb{Q}(\omega^I, B^*)$ is $\mathcal{F}^I$-measurable $\forall B^* \in \mathcal{F}^*$, we use

$$\mathbb{Q}(\omega^I, f) = \prod_{j=1}^{J} \exp\left[\int_{t_{j-1}}^{t_j} \left(u_j \gamma(s)(\omega^I) - \frac{u_j^2 \sigma^2(s)(\omega^I)}{2} + \right.\right.$$

$$\left.\left. \int_{\mathbb{R}} \left(e^{iu_j x} - 1 - iu_j x 1_{|x|\leqslant 1}\right) \mu(s, dx)(\omega^I)\right) ds\right], \ f \in \mathcal{G}^*,$$

where $\mathcal{G}^* = \{f : f = \prod_{j=1}^{J} e^{iu_j(X^*(t_j) - X^*(t_{j-1}))}, \ u_j \in \mathbb{R}, \ 0 = t_0 < \ldots < t_J < \infty, \ J \in \mathbb{N}\}$. Since γ, σ^2 and μ are progressively measurable, $\omega^I \mapsto \mathbb{Q}(\omega^I, f)$ is $\mathcal{F}^I$-measurable for all $f \in \mathcal{G}^*$. As $\mathcal{G}^*$ generates $\mathcal{F}^*$, $\omega^I \mapsto \mathbb{Q}(\omega^I, B^*)$ is $\mathcal{F}^I$-measurable $\forall B^* \in \mathcal{F}^*$ (Klenke (2008)[Bemerkung 8.25]). □

Theorem 4.1 shows that we can construct a process X from stochastic local characteristics. However, without conditioning on a path $I(\omega)$, we do not know which class of processes X is assigned to. In the following we will show that X is a semimartingale.

Semimartingales (see Jacod and Shiryaev (2002) for any unexplained but classical notation) represent an important class of stochastic processes. Since they frequently occur as solutions to stochastic differential equations, they can be found indirectly in many applications, especially in finance. In contrast to Lévy processes, however, general semimartingales are rarely in the focus of statistical analysis. This is due to the fact that they are analytically more difficult to handle, which can make estimation and simulation difficult or even impossible. Fortunately, in our case the situation is different as we will see later.

Theorem 4.2
Assume X is an I-CII process on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ as in Theorem 4.1. Then, using the filtration $\mathbb{F} = \{\mathcal{F}_t^I \otimes \mathcal{F}_t^X\}_{t\geqslant 0}$, where $\mathcal{F}_t^X = \sigma(X(s) : s \leqslant t)$, X is a semimartingale. Furthermore, if for some constant $C > 0$,

$$\left|\mathbb{E}\left(X(t_i) - X(t_{i-1})|I\right)\right| < C, \ 0 \leqslant t_{i-1} < t_i < \infty,$$

then X is an Itô-semimartingale with spot characteristics (γ, σ^2, μ).

Proof. Define

$$\begin{aligned} N(t) =& X(t) - \sum_{s \leqslant t} \Delta X(s) 1_{|\Delta X(s)|>1} - \int_0^t \gamma(s) ds, \\ N^g(t) =& \sum_{s \leqslant t} g(\Delta X(s)) - \int_0^t \int_{\mathbb{R}} g(x) \mu(s, dx) ds, \text{ and} \\ N^*(t) =& N^2(t) - \int_0^t \sigma^2(s) ds - \int_0^t \int_{\mathbb{R}} x^2 1_{|x| \leqslant 1} \mu(s, dx) ds, \end{aligned}$$

where g is a bounded function with $g(x) = 0$, $|x| < \epsilon$, $\epsilon > 0$. If $N(t)$ is a local martingale, we see immediately, that X is a semimartingale.

Let us denote its characteristics by (B, C, q), then $B = \int_0^t \gamma(s) ds$. If also $N^g(t)$ is a local martingale, $q(ds, dx) = \mu(s, dx) dx$, by Jacod and Shiryaev (2002)[Theorem II.1.8]. To find the characteristic $C = \langle X^c, X^c \rangle$, we use that for each locally square-integrable martingale N, $\langle N, N \rangle$ is the unique predictable process such that $N^2 - \langle N, N \rangle$ is a local martingale. Since $\langle N, N \rangle$ is the compensator of $[N, N] = \langle N^c, N^c \rangle + \sum_s (\Delta N(s))^2$ (Jacod and Shiryaev (2002)[Proposition I.4.50 and Theorem I.4.52]), if $N^*(t)$ is a local martingale, $C(t) = \int_0^t \sigma^2 ds$. It is obvious that X is then also an Itô-semimartingale with spot characteristic (γ, σ^2, μ).

It remains to be shown that N, N^g and N^* are locale martingales. Since X has I-conditionally independent increments, this is true under the conditional probability measure $\mathbb{P}(X \in \cdot | \mathcal{F}^I)(\omega)$ for each $\omega \in \Omega$. For now, consider the filtration $\mathbb{F}^* = \{\mathcal{F}^I \otimes \mathcal{F}_t^X\}_{t \geqslant 0}$. Then, we can use the fact that for each stochastic process Y, it holds

$$\begin{gathered} \forall \omega \in \Omega : Y \text{ is a martingale with respect to } \mathbb{F}^* \text{ and } \mathbb{P}(\cdot | \mathcal{F}^I)(\omega) \\ \Rightarrow \\ Y \text{ is a martingale with respect to } \mathbb{F}^* \text{ and } \mathbb{P}, \end{gathered}$$

which is easy to show. By a simple localization this shows that N, N^g and N^* are local martingales with respect to $\mathbb{F}^*$ and $\mathbb{P}$. Hence, X is a semimartingale for the filtration $\mathbb{F}^*$.

However, since X is also $\mathbb{F}$-adapted, it is a $\mathbb{F}$ semimartingale (Protter (2005)[Theorem II.4]). Furthermore, if the $\mathcal{F}^I$-conditional expectation of the increments of X is bounded, by Föllmer and Protter (2011)[Corollary 3.1], the processes N, N^g and N^* remain local martingales with respect to $\mathbb{F}$. This finishes the proof. □

The property that X is a semimartingal allows us to define stochastic integrals with respect to X. This way we can construct more complex stochastic models in which the stochastic behavior is determined by a CII process (see Section 4.2). Also a procedure

for simulation of semimartingales is given by Jacod and Protter (2012)[Section 5.6.3].

4.1.2 Nonparametric estimation

Estimation for general semimartingales can be difficult or even impossible. Fortunately, in the case of an I-CII process, using the independence of the increments under $\mathbb{P}(\cdot|\mathcal{F}^I)$, the situation is different. We will use this property to show that for a set of observations $\{(X(t_i), I(t_i))\}_{i=0,1,\ldots,n}$, $0 = t_0 < t_1 < \ldots < t_n = n\Delta = T$, where $|t_i - t_{i-1}| = \Delta > 0$, a consistent ($T \to \infty$) estimator of the moments $\mathbb{E}\left((\Delta_i X)^k | I(t_{i-1}) = x\right)$, $\Delta_i X := X(t_i) - X(t_{i-1})$, exists. For this we need the following assumption to be fulfilled.

Assumption 4.3
Each component of the I-adapted triplet (γ, σ^2, μ) is a piecewise constant function of I, i.e. for $t \in [t_i, t_{i+1}]$,

$$\gamma(t) = \gamma(t_i),\ \sigma^2(t) = \sigma^2(t_i),\ \mu(t, B) = \mu(t_i, B),\ B \in \mathcal{B}(\mathbb{R}),$$

and depends on a deterministic functions $\bar{\gamma}$, $\bar{\sigma}^2$ resp. a family of measures $\{\bar{\mu}_x\}_{x\in\mathbb{R}}$, such that

$$\gamma(t_i) = \bar{\gamma}(I(t_i)),\ \sigma^2(t_i) = \bar{\sigma}^2(I(t_i)),\ \mu(t_i, B) = \bar{\mu}_{I(t_i)}(B),\ B \in \mathcal{B}(\mathbb{R}).$$

If Assumption 4.3 is satisfied a family of infinitely divisible distributions $\{H_x\}_{x\in\mathbb{R}}$, connected to the I-CII process X by the following remark, can be defined, where each distribution H_x has a generating triplet $(\bar{\gamma}(x), \bar{\sigma}^2(x), \bar{\mu}_x)$.

Remark 4.4
If Assumption 4.3 is satisfied, then, the increments $\Delta_i X | I(t_{i-1}) = x \sim \Delta_i L$, where $\{L(t)\}_{t\geqslant 0}$ is a Lévy process with $L(1) \sim H_x$. Since H_x is usually chosen such that statistical properties are easy to handle, this allows us to study the properties of $\Delta_i X$ in a simple way.

If H_x is furthermore a parametric distribution, i.e. there exists a parameter function $\theta : \mathbb{R} \to \Theta$ with parameter space Θ such that $(\bar{\gamma}(x), \bar{\sigma}^2(x), \bar{\mu}_x) = f(\theta(x))$ for a function f, and allows the use of the method of moments, we can determine $\theta(x)$ from the estimates of

$$m_k(x) := \mathbb{E}\left((\Delta_i X)^k | I(t_{i-1}) = x\right),\ k \in \mathbb{N}. \tag{4.2}$$

This gives us a nonparametric estimator of θ.

Since H_x is the distribution of the increments $\Delta_i X$ for $I(t_{i-1}) = x$, estimation at x based on the increments of X only makes sense if I spends enough time around x. This was also report for nonparametric estimation of the diffusion coefficient for an one-dimensional diffusion process by Jacod (2000). Equal to the process used there, to allow easy measurement of this time, we assume that $\{I(t)\}_{t\geqslant 0}$ is a continuous adapted process of the form

$$I(t) = I_0 + \int_0^t b^I(s)ds + \int_0^t \sigma^I(I(t))dW^I(s), \tag{4.3}$$

where $\{W^I(t)\}_{t\geqslant 0}$ is a standard Brownian motion, $b^I \in C_b^2(\mathbb{R})$ and $\sigma^I \in C_b^3(\mathbb{R})$ and nonvanishing. Then, as we will see later, the time I spends around x goes to infinity as T goes to infinity.

Remark 4.5
Note that unlike Florens-Zmirou (1993) and Jacod (2000), we consider the process X for fixed Δ and $T \to \infty$. So we are not in a high-frequency setting, i.e. we do not consider infill asymptotics $\Delta \to 0$. The estimate of $m_k(x)$ is therefore not limited to x for which we have enough observations $I(t_i) \approx x$.

To estimate the conditional expectation $m_k(x)$ as defined in (4.2), inspired by the Nadaraya-Watson estimator, we use

$$\widehat{m}_{k,n}(x) := \frac{\sum_{i=1}^n K_{\epsilon_n}(I(t_i) - x)\,(\Delta_i X)^k}{\sum_{i=1}^n K_{\epsilon_n}(I(t_i) - x)}, \tag{4.4}$$

where $K_h(x) = K(\frac{x}{h})$, $h > 0$, for a bounded, positive, symmetrical function $K : \mathbb{R} \to \mathbb{R}^+$ such that $\int K(x)dx = 1$. The denominator of $\widehat{m}_{k,n}(x)$ is also referred to as $N_n(x)$.

Theorem 4.6
For $k, x \in \mathbb{R}$ assume $m_j(y) \in C^1(B_\epsilon(x))$, $j \in \{k, 2k\}$, $B_\epsilon(x) := \{y : |x - y| \leqslant \epsilon\}$, $\epsilon > 0$, the $4k$th moment of H_y is bounded and $m_{2k}(y) - m_k^2(y)$ is non-vanishing for $y \in B_\epsilon(x)$. Let $A \subset \mathcal{F}$ with $\mathbb{P}(A) > 0$ such that $N_n(x) \to \infty$ and $N_n(x)\epsilon_n^2 \to 0$ on A. Then,

$$\sqrt{N_n(x)}\,(\widehat{m}_{k,n}(x) - m_k(x)) \xrightarrow[n\to\infty]{\mathcal{L}_{|A}} \mathcal{N}\left(0, m_{2k}(x) - m_k^2(x)\right),$$

where for each random variable Y the notation $\mathcal{L}_{|A}(Y) = \mathcal{L}(Y|A)$ stands for the law of Y under the conditional probability $\mathbb{P}_{|A} = \mathbb{P}(\cdot|A)$.

Proof. W.l.o.g. we only consider the case $K_\epsilon(y - x) = \mathbf{1}_{\{y\in B_\epsilon(x)\}}$. To prove the statement, we first consider the estimator $\widehat{m}_{k,n}(x)$ conditionally on a path of I such that the above

conditions are satisfied. Then the increments $\Delta_i X$, $i = 1, \dots, n$, are independent. This allows to apply a central limit theorem, which requires to normalize first:

$$\frac{N_n(x)}{\sqrt{\sum_{i=1}^n s_k(I(t_i))\mathbf{1}_{\{I(t_i)\in B_{\epsilon_n}(x)\}}}}\widehat{m}_{k,n}(x) =$$

$$\frac{\sum_{i=1}^n \left((\Delta_i X)^k - m_k(I(t_i))\right)\mathbf{1}_{\{I(t_i)\in B_{\epsilon_n}(x)\}}}{\sqrt{\sum_{i=1}^n s_k(I(t_i))\mathbf{1}_{\{I(t_i)\in B_{\epsilon_n}(x)\}}}} + \frac{\sum_{i=1}^n m_k(I(t_i))\mathbf{1}_{\{I(t_i)\in B_{\epsilon_n}(x)\}}}{\sqrt{\sum_{i=1}^n s_k(I(t_i))\mathbf{1}_{\{I(t_i)\in B_{\epsilon_n}(x)\}}}}, \tag{4.5}$$

where $s_k(y) := m_{2k}(y) - m_k^2(y) > 0$ for $y \in B_{\epsilon_n}(x)$, $\epsilon_n < \epsilon$.

By the boundedness of the $4k$th moment of H_y for $y \in B_{\epsilon_n}(x)$ and $N_n(x) \to \infty$ for the given path of I, Lyapunov's CLT can be applied to the first summand in (4.5). Considering the second summand, we have

$$\sum_{i=1}^n \left(m_k(I(t_i)) - m_k(x)\right)\mathbf{1}_{\{I(t_i)\in B_{\epsilon_n}(x)\}} \in \mathcal{O}\left(\epsilon_n N_n(x)\right), \tag{4.6}$$

as $m_k(y) = m_k(x) + \mathcal{O}(\epsilon_n)$, $y \in B_{\epsilon_n}(x)$. Hence,

$$\sqrt{N_n(x)}\left(\widehat{m}_{k,n}(x) - m_k(x)\right) \sim \sqrt{\frac{\sum_{i=1}^n s_k(I(t_i))\mathbf{1}_{\{I(t_i)\in B_\epsilon(x)\}}}{N_n(x)}}Z + \mathcal{O}\left(\epsilon_n\sqrt{N_n(x)}\right),$$

for $Z \sim \mathcal{N}(0,1)$. By the same argument as in (4.6),

$$\frac{\sum_{i=1}^n s_k(I(t_i))\mathbf{1}_{\{I(t_i)\in B_\epsilon(x)\}}}{N_n(x)} = s_k(x) + \mathcal{O}(\epsilon_n).$$

Since $\epsilon_n \to 0$ and $\epsilon_n\sqrt{N_n(x)} \to 0$, applying Slutsky's theorem,

$$\sqrt{N_n(x)}\left(\widehat{m}_{k,n}(x) - m_k(x)\right) \xrightarrow[n\to\infty]{\mathcal{L}_{|I}} \mathcal{N}\left(0, m_{2k}(x) - m_k^2(x)\right),$$

where this limit is valid given a specific appropriate path I. However, since the limiting law does not depend on I, we have the same under $\mathbb{P}$. □

Theorem 4.6 ensures that for each suitable sequence of bandwidths ϵ_n we asymptotically get a proper estimator of $m_k(x)$ at each x. However, for real application an appropriate choice of $\epsilon = \epsilon_n$ plays an important role. Since we usually want to get a smooth estimator of m_k, we would like to have a bandwidth ϵ^* such that

$$\epsilon^* = \operatorname*{argmin}_{\epsilon>0} \int_{\mathbb{R}} \left(m_k(x) - \widehat{m}_k^\epsilon(x)\right)^2 dx,$$

where $\widehat{m}_k^\epsilon$ corresponds to $\widehat{m}_{k,n}$ based on a bandwidth of ϵ.

In approximation to the upper integral, we choose the bandwidth ϵ according to

$$\min_{\epsilon>0} \sum_{i=1}^{n} \left((\Delta_i X)^k - \widehat{m}_{k,-i}^\epsilon(I(t_i)) \right)^2, \tag{4.7}$$

where $\widehat{m}_{k,-j}^\epsilon(x)$ is the estimator of $m_k(x)$ based on the observations $\{(X(t_i), I(t_i))\}_{i\neq j}$. This approach is similar to the least squares cross-validation for estimation of the smoothing parameter in nonparametric regression (cf. Li and Racine (2007)[2.2.2]).

4.1.3 Testing

To check, whether a model with variable parameters is really necessary, we need to test against the alternative of a stochastic process with constant characteristics, i.e. a Lévy process. Therefore we consider the moments $m_k(x)$, $k \in \mathbb{N}$, and test,

$$H_0 : \forall x \; m_k(x) = m_k \quad \text{against} \quad H_1 : \exists x \; m_k(x) \neq m_k,$$

for some unknown constant m_k. An intuitive test statistic would be given through

$$\frac{1}{\bar{m}_{2k} - \bar{m}_k^2} \int_{\mathbb{R}} N_n(x) \left(\widehat{m}_k(x) - \bar{m}_k \right)^2 dx,$$

where $\bar{m}_k := \frac{1}{n} \sum_{i=1}^{n} (\Delta_i X)^k$. Here, $N_n(x)$ serves as a weighting function and also simplifies the asymptotic analysis, since it is the proper scaling component by Theorem 4.6.

For this test, however, the asymptotic distribution is difficult to derive because the estimators $\widehat{m}_k(x)$ for different x are generally not independent of each other. We therefore use a discretized version of the upper integral, namely,

$$S_n := \frac{1}{\bar{m}_{2k} - \bar{m}_k^2} \sum_{i=1}^{J} N_n(x_j) \left(\widehat{m}_k(x_j) - \bar{m}_k \right)^2, \tag{4.8}$$

for $x_1, x_2, \ldots, x_J \in \mathbb{R}$, $J \in \mathbb{N}$, with $|x_i - x_j| > c\epsilon_n, \forall i, j$, $c > 2$. Here, using a kernel K with support $[-1, 1]$, the estimators $\widehat{m}_k(x_j)$ are based on non-overlapping intervals and therefore are independent given I.

Theorem 4.7

Assume K has support $[-1, 1]$. If the assumptions of Theorem 4.6 are satisfied, then

$$S_n \xrightarrow[n\to\infty]{\mathcal{L}_{|A}} \chi_J^2,$$

where for each random variable Y the notation $\mathcal{L}_{|A}(Y) = \mathcal{L}(Y|A)$ stands for the law of Y under the conditional probability $\mathbb{P}_{|A} = \mathbb{P}(\cdot|A)$.

Proof. As $\widehat{m}_k(x_j)$, $j = 1, 2, \ldots, J$, are independent, w.l.o.g. it is enough to consider the case $J = 1$. By

$$n\bar{m}_k = N_n(x)\widehat{m}_k(x) + N_n^c(x)\widehat{m}_k^c(x),$$

where $N_n^c(x) := \sum_{i=1}^n \mathbf{1}_{\{I(t_i)\in B_{\epsilon_n}^c(x)\}}$, $\widehat{m}_k^c(x) := \frac{1}{N_n^c(x)} \sum_{i=1}^n (\Delta_i X)^k \mathbf{1}_{\{I(t_i)\in B_{\epsilon_n}^c(x)\}}$ for $B_{\epsilon_n}^c(x) = \{y : |y - x| > \epsilon_n\}$, we get

$$\sqrt{N_n(x)}\,(\widehat{m}_k(x) - \bar{m}_k) = \left(1 - \frac{N_n(x)}{n}\right)\sqrt{N_n(x)}\widehat{m}_k(x) - \frac{N_n^c(x)}{n}\frac{\sqrt{N_n(x)}}{\sqrt{N_n^c(x)}}\sqrt{N_n^c(x)}\widehat{m}_k^c(x).$$

First, by Theorem 4.6, we have under H_0,

$$Z_n := \sqrt{N_n(x)}(\widehat{m}_k(x) - m_k) \xrightarrow{\mathcal{L}_{|A}} Z \text{ and } Z_n^c := \sqrt{N_n^c(x)}(\widehat{m}_k^c(x) - m_k) \xrightarrow{\mathcal{L}_{|A}} Z^c,$$

for two independent random variables Z, Z^c with $Z, Z^c \sim \mathcal{N}(0, m_{2k} - m_k^2)$. Moreover, under H_0, using the law of large numbers, for an appropriate path of I, $\bar{m}_k \xrightarrow{\mathbb{P}_{|I}} m_k$, where m_k does not depend on I. Hence,

$$V_n := \frac{1}{\sqrt{\bar{m}_{2k} - \bar{m}_k^2}} \xrightarrow{\mathbb{P}_{|A}} \frac{1}{\sqrt{m_{2k} - m_k^2}} =: V.$$

Now, using Slutsky's theorem, we get

$$V_n Z_n \overset{\mathcal{L}_{|A}}{\to} VZ \sim \mathcal{N}(0, 1) \text{ and } V_n Z_n^c \overset{\mathcal{L}_{|A}}{\to} VZ^c \sim \mathcal{N}(0, 1).$$

To determine the limit of $N_n(x)$, we use the occupation time formula for semimartingales (see Protter (2005)[Chapter IV.5, Corollary 1]), which yields

$$\int_0^{n\Delta} \mathbf{1}_{\{I(s)\in B_{\epsilon_n}(x)\}}\sigma^2(s)ds = \int_{B_{\epsilon_n}(x)} L^a(s)da \in \mathcal{O}(n\Delta\epsilon_n),$$

by $L^a(s) \leqslant n\Delta$ a.s.. So, by the boundedness of σ and $\epsilon_n \to 0$, we obtain $\frac{N_n(x)}{n} \overset{a.s.}{\to} 0$. Since $\frac{N_n(x)+N_n^c(x)}{n} = 1$, we also have $\frac{N_n^c(x)}{n} \overset{a.s.}{\to} 1$ and $\frac{N_n(x)}{N_n^c(x)} \overset{a.s.}{\to} 0$. Now, we can use Slutsky's theorem again to finish the proof. □

When applying the test in Theorem 4.7, the question arises how many evaluation points x_i, $i \in J$, should be chosen. For a high explanatory power of the test a large number of evaluation points seems reasonable, where the condition $|x_i - x_j| > c\epsilon_n$, $\forall i, j$, $c > 2$, results

in an upper limit given a bandwidth ϵ_n. However, using a bandwidth that smoothes the estimator as in (4.7) is not appropriate under the null.

Instead, the bandwidth ϵ_n can be considered as a tuning parameter that is chosen in such a way that the test has a high quality, i.e. that the empirical rejections rates match the theoretical ones. The empirical rejection rates for different bandwidths can be found in simulation studies.

4.1.4 Numerical results

To verify the quality of the estimator introduced in Subsection 4.1.2 and the test introduced in Subsection 4.1.3, we perform a simulation study. In accordance with our application in Section 4.2, we assume that the distribution of the increments of the I-CII process X corresponds to a NIG distribution. As external process I, we use the residual demand process R presented there. Therefore, our simulation study is based on $n = 17544$ observations.

First, we want to evaluate the performance of our estimation procedure. The method consists of applying the method of moments based on the nonparametric estimators $\hat{m}_{k,n}$ (cf. Subsection 4.1.2). To this end we assume that the true parameter functions α, β, μ and δ are the same as in our application (see Figure 4.8) and look at the error

$$e_g = \mathbb{E}\left[\left.\frac{\int_U \left(\hat{g}(x) - g(x)\right)^2 dx}{\int_U g^2(x)dx}\right| \mathcal{F}^I\right], \ U \subset \mathbb{R},$$

for the function g to be estimated. For this conditional expectation, we only consider the case, where I is the observed residual demand process, since we do not want to let errors, which result from an inaccurate modeling of the external process, distort our results. The interval U is chosen so that there are enough observations $I(t_i) \approx x$ for each $x \in U$. In the following we use the same interval as for the estimation of the parameter functions in Section 4.2, i.e. $U = [14.7, 66.3]$. Then the error is approximated by 10000 simulations at a sample variance of not larger than 0.5, where the integrals are evaluated on a sufficiently fine grid. Since for $|\chi| \approx \xi \approx 1$ the problem of choosing α and β is ill-conditioned (see Remark 4.9), we removed 478 simulations, where $|\chi| > 0.95$.

The error based on the remaining simulations for each parameter function and the related skewness χ and steepness ξ are given in Table 4.1. We see that the estimation procedure provides satisfactory results.

Next, we want to assess the quality of testing for a CII process as proposed in Theorem 4.7.

Table 4.1: Performance of the nonparametric parameter estimators.

g	α	β	μ	δ	χ	ξ
$\widehat{e}_g$	0.25	0.4	0.41	0.24	0.19	0.002

To this end, we first use the 10000 paths of X for which the parameter estimation was performed above, and assess the rejection rates under the alternative, i.e. assuming that X depends on I. Hence, we check our test procedure under the same circumstances as in our empirical application.

In particular, we assume again that the increments are NIG distributed with possibly changing parameters. Since the parameters of a NIG distribution can be determined by its first four moments, we apply the test to m_k, $k = 1, 2, 3, 4$, for each simulation. We use a uniform kernel K with bandwidth ϵ selected as described at the end of Subsection 4.1.3. Figure 4.1 shows the empirical rejection rates. In this case, the null is rejected with high probability, in particular for the first and second moment.

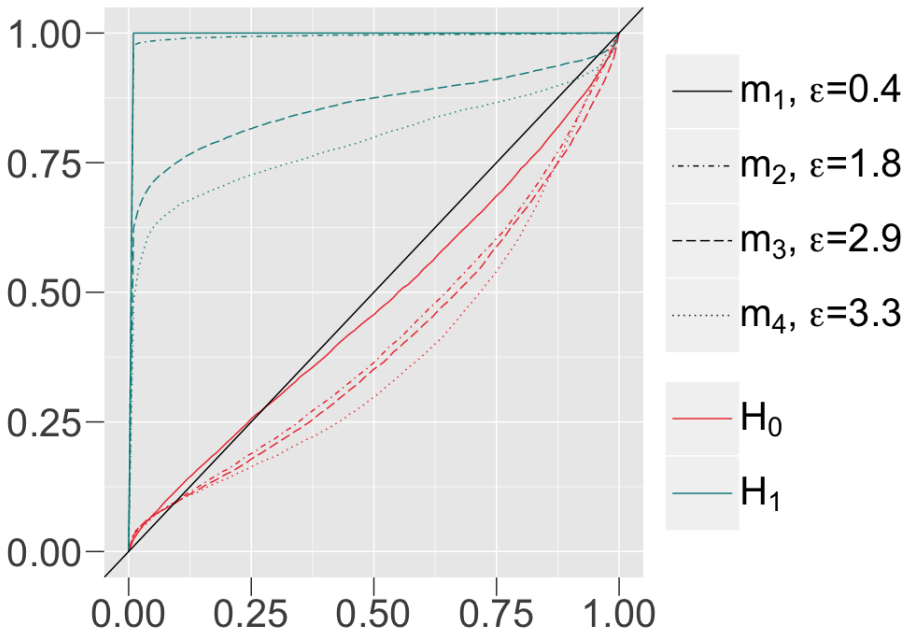

Figure 4.1: Empirical rejection curves for the test from Theorem 4.7 and used bandwidth ϵ.

Second, we simulate 10000 paths of a process X^* under the null, i.e. under the hypothesis that X^* is actually a Lévy process. Again, the parameters of X^* are chosen according to our empirical application, but with the assumption of constant parameters, by fitting an NIG distribution to the increments shown in Figure 4.5. The corresponding method of moments estimates are listed in Table 4.2. Especially for the usual significance

Table 4.2: Results for fitting a NIG-Lévy process to the process X of Subsection 4.2.3.

α	β	μ	δ	χ	ξ
0.014	0	0.04	0.56	0.99	-0.02

levels of up to 10% they are quite close to their theoretical values.

We conclude that the test is very reliable, at least in the circumstances of our application (and at usual significance levels).

4.2 Application to the intraday market

In the following we show how electricity price variations can properly described by a CII process using information on demand and renewable power generation. We use the EPEX SPOT Intraday index (ID_3-*Price*) of the hourly contracts on the intraday market as price reference (cf. Subsection 1.3.4). To quantify the effect of renewable energies on the ID_3-Price, we follow Wagner (2012) (cf. Subsection 1.4.3) and use residual demand as external factor. Residual demand is defined as total load minus infeed from renewables. Renewable power is produced at vanishing marginal costs. Hence, they can be assumed to reduce demand according to their production level.

Among the renewable energies we focus on solar and wind power, since they have the largest installed capacities in Germany and are responsible for most of the variability in the electricity supply. Our study is based on hourly data for the period Jan 01, 2016 to Dec 31, 2017. All data are provided by the European Energy Exchange (EEX).

4.2.1 A structural price model

Models that explain the electricity price as a function of physical or economic factors are called structural models (cf. Section 1.4). In the following we will use a structural model similar to the one defined by Burger et al. (2004).

Assumption 4.8
The electricity price process $S = \{S(t)\}_{t\geqslant 0}$ is given by

$$S(t) = f(R(t)) + Y(t) + Z(t),\ t \geqslant 0,$$

where f is a deterministic supply function whose input are the realizations of the real-valued stochastic process $R = \{R(t)\}_{t\geqslant 0}$. $Y = \{Y(t)\}_{t\geqslant 0}$ and $Z = \{Z(t)\}_{t\geqslant 0}$ are real-valued

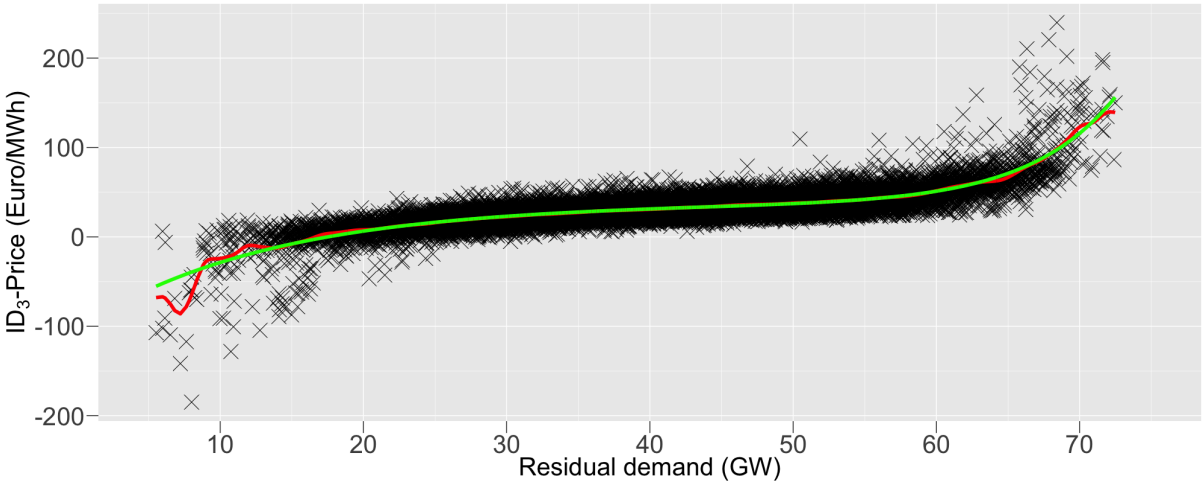

Figure 4.2: ID_3-Price S in dependence to residual demand R with fitted supply-function f based on nonparametric regression (red) and the dual exponential model (green).

stochastic processes with $Y(0) = 0$ *while* Z *has zero mean. For the factors* R, Y *and* Z*, we have* $R \perp\!\!\!\perp Y$ *and* $Y \perp\!\!\!\perp Z$*, where* $\perp\!\!\!\perp$ *denotes independence of random variables.*

In Assumption 4.8, R models the residual demand and so f refers to the price charged by conventional producers. Therefore, f is called the supply function. Y accounts for the non-stationary long-term variation of the electricity price. Such variations can be caused, for example, by changed fuel prices. In contrast, Z accounts for stationary short-term variation, which can be not explained by f. An appropriate dynamic and functional form of each model component is given in the following.

While it is also possible to use an availability-adjusted residual demand in Assumption 4.8, we found that this does not improve the model, a result which is in line with Wagner (2012).

4.2.2 The supply function

Figure 4.2 shows the electricity price in dependence on residual demand. As expected, the price of electricity will rise as demand increases. The link seems strictly nonlinear. A parametric form of f is given by the *dual exponential model* of Yang et al. (2013), which used

$$f(x) = a + \frac{\exp\left(\frac{x-b}{c}\right) - \exp\left(-\frac{x-d}{e}\right)}{2},$$

as the supply function for the EEX market. The estimated parameters using the least-squares method are shown in Table 4.3. The fitted supply function seems reasonable also compared with the result of a nonparametric regression (see Figure 4.2). Nevertheless,

we can see that especially extreme prices are not sufficiently described by the supply function (see also Figure 4.3). The remaining variation will therefore be explained by Y and Z.

Table 4.3: Least-squares estimators of the dual exponential supply function.

a	b	c	d	e
37.61	39.49	6.02	74.85	13.27

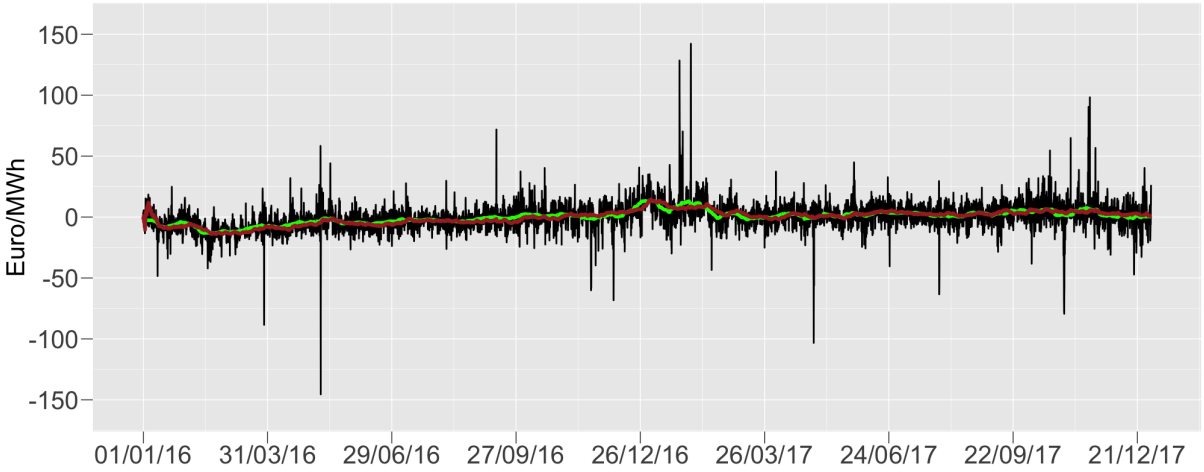

Figure 4.3: Residual price $S - f(R)$ (black) and moving average (green) resp. Kalman filter (brown) of Y.

4.2.3 The short-term variation

We start with the definition of the short-term variation Z. Beside some possible non-stationary behavior, which is to be modeled by Y, Figure 4.3 show that there is a volatile component of the residual price $S - f(R)$, that returns to the mean-level very fast. This short-term price variation can be caused e.g. by unexpected weather phenomena, which impact the infeed from renewable energies.

To make jumps and mean-reversion possible, we assume that Z follows the stochastic differential equation

$$dZ(t) = -\lambda(Z(t))Z(t)dt + dX(t),\ t \geqslant 0,\ Z_0 = 0, \tag{4.9}$$

where $\lambda \in C_b^1(\mathbb{R}^+)$, $X = \{X(t)\}_{t \geqslant 0}$ is a zero mean semimartingale. Note, that Equation (4.9) admits a unique strong solution with zero mean without any restriction on X (see

e.g. Protter (2005) [Chapter V.3]). The process Z belongs to the class of *generalized Ornstein-Uhlenbeck processes*, cf. Maller et al. (2009). Since the variation of Z should depend on R (see Figure 4.2), X will be modeled as a R-CII process. The reason to use a mean-reversion function $\lambda : \mathbb{R} \to \mathbb{R}^+$ instead of a constant is, that after a jump Z should be allowed to return to the mean level much faster than usual. Such a behavior of the price process is often observed in electricity markets (cf. Chapter 2).

For calibration of Z, we are faced with the problem, that the observations $S(t_i) - f(R(t_i))$, $0 = t_0 < t_1 < \cdots < t_n = n\Delta = T$, where $|t_i - t_{i-1}| = \Delta$, depend on $Z(t_i)$ and on $Y(t_i)$. However, since Z has a mean level of zero and Y models the low-frequency price dynamics, we can use a moving average filter to get a first approximation $\widehat{Y}$ of the long-term variation. The filter with window size of two weeks is shown in Figure 4.3. Now, we use the values $Z(t_i) = S(t_i) - f(R(t_i)) - \widehat{Y}(t_i)$ for calibration of Z.

An approximate model for Z is given by the *Euler scheme*

$$\Delta_i Z = -\lambda(Z(t_{i-1}))Z(t_{i-1})\Delta + \epsilon_i^Z, \tag{4.10}$$

where $\epsilon_i^Z = \Delta_i X$. Since X is assumed to be a zero mean R-CII process, given the observations $R(t_i)$, $\{\epsilon_i^Z\}_{i=1,\dots,n} \sim \mathcal{WN}$, where ϵ_i^Z has mean 0 and variance depending on $R(t_{i-1})$ for all i. For small Δ, this approximation is sufficiently accurate (see Jacod and Protter (2012) [Section 5.6.3]).

Now, to get an estimator for λ, we distinguish between two price regimes. First, a mean-reversion rate λ_{base} for regular price levels is found by the method of least-squares, i.e.

$$\lambda_{\text{base}} = \min_{\lambda>0} \sum_{i \in \mathcal{T}_{\text{base}}} \left(\Delta_i Z + \lambda Z(t_{i-1})\Delta\right)^2, \tag{4.11}$$

where $\mathcal{T}_{\text{base}} = \{2 \leqslant i \leqslant n : |Z(t_{i-1})| \leqslant 3 \cdot \sigma_Z, |Z(t_i)| \leqslant 3 \cdot \sigma_Z\}$ for σ_Z being the empirical standard deviation of $\{Z(t_i)\}_{i=0,\dots,n}$.

Next, to get a mean-reversion rate λ_{spike} for extreme price data, we calculate the least-squares estimator as in (4.11), but for the observations $\mathcal{T}_{\text{spike}} = \{2 \leqslant i \leqslant n : |\Delta_i Z| > 3 \cdot \sigma_{\Delta Z}, \Delta_i Z \Delta_{i+1} Z < 0\}$, where $\sigma_{\Delta Z}$ is the empirical standard deviation of $\{\Delta_i Z\}_{i=1,\dots,n}$. For small Δ, λ_{spike} can be seen as the instantaneous slope of $t \to Z(t)$ right after a jump. Here, the supplementary condition $\Delta_{i-1} Z \Delta_i Z < 0$ is used to correctly identify a jump and not to mistake it by a large increments caused by a strong mean-reversion. A similar estimator for a mean-reverting compound Poisson process was used by Deschatre (2017). The results are shown in Table 4.4, where we used $\Delta = 1$ as in the following.

With this we define λ as the function which takes the value λ_{base} for $|x| \leqslant 3 \cdot \sigma_Z - \delta$, λ_{spike} for $|x| \geqslant 3 \cdot \sigma_Z + \delta$ and is appropriately interpolated in between for some small $\delta > 0$.

Table 4.4: Parameters of the mean-reversion function of Z.

σ_Z	λ_{base}	λ_{spike}
10	0.13	0.28

Filtering the CII process Now, to specify Z, it remains to calibrate the R-CII process X. To do so, we write (4.9) as

$$\int_{t_{i-1}}^{t_i} dX(t) = \int_{t_{i-1}}^{t_i} dZ(t) + \int_{t_{i-1}}^{t_i} \lambda(Z(t))Z(t)dt,$$

from which we get

$$\Delta_i X \approx \Delta_i Z + \frac{\lambda(Z(t_i))Z(t_i) + \lambda(Z(t_{i-1}))Z(t_{i-1})}{\Delta/2}. \tag{4.12}$$

Figure 4.4 shows $\{\Delta_i X\}_{i=1,\dots,n}$ in dependence on residual demand. We see that price jumps are increasing both in frequency and in magnitude towards the boundaries. This justifies the modeling of X as R-CII process. What is surprising at first glance is that

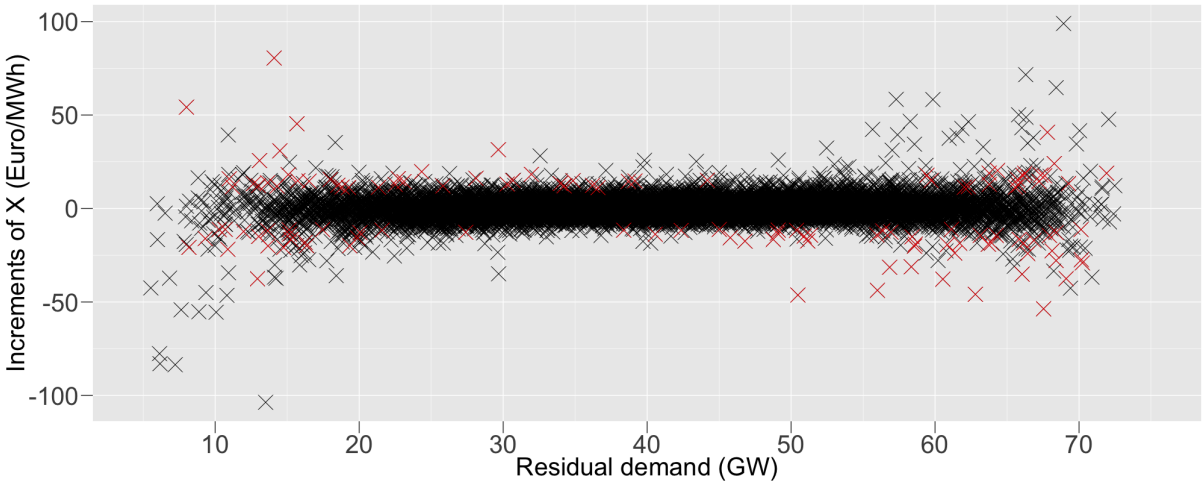

Figure 4.4: Increments of X (black) and removed jumps (red).

the sign of the jumps is not connected to the side of the boundary. This contradicts our expectations (see Figure 4.2). The reason is that actually the mean-reversion rate is quite individual and so some increments $\Delta_i X$ derived by (4.12) contain a significant mean-reversion part, especially in the case of a previous large price jump. We use a

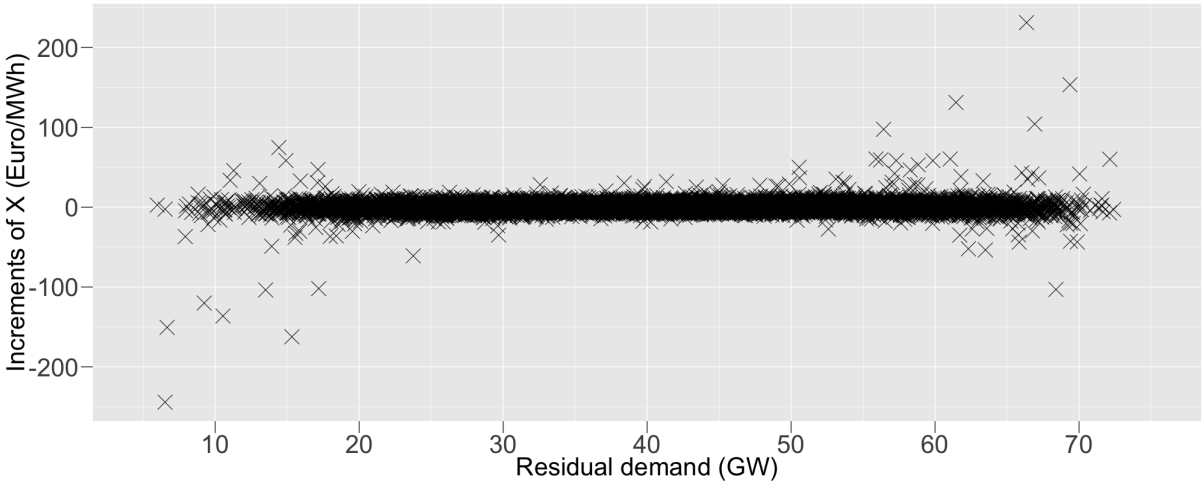

Figure 4.5: Preprocessed increments of X.

statistical approach to overcome this.

Data preprocessing for the CII process To remove the part depending on Z contained in the increments $\{\Delta_i X\}_{i=1,\dots,n}$, we first identify the jumps of X. We classify a price movement as a jump if it exceeds the bound $\pm\Theta \cdot \sigma_X$, where $\Theta > 0$ and σ_X is the empirical standard deviation of $\{\Delta_i X\}_{i=1,\dots,n}$. This is performed iteratively, where in each step the jumps are removed and σ_X is calculated based on the remaining increments. The algorithm stops if no further jump is detected. The threshold level Θ is usually chosen between 2 and 3 (cf. Weron (2006)).

Next, we remove all jumps from the data that were preceded by a jump of opposite sign within a time of $T_{nojump} > 0$. For the choice of T_{nojump} we suggest to use some data driven criterion. Using a threshold level $\Theta = 3$, we set $T_{nojump} = 5$ since the fact that a jump is no longer visible after about 5 hours seems to be consistent to the intraday behavior of electricity prices (see also Figure 4.6).

The removed increments are highlighted in the Figure 4.4. Now X behaves as we would expect it to. We almost exclusively observe negative jumps in case of low residual demand, whereas positive jumps occur more frequently for high residual demand.

Figure 4.6 shows another issue in calibrating X based on the filtered increments $\{\Delta_i X\}_{i=1,\dots,n}$. Extreme prices are often not the result of a single big jump but of a series of smaller movements. Although the dependence of the jump intensity of X on R allows clustering of jumps, this is not as strong as the data suggests. Hence, to generate prices of appropriate size by our model, we add up jumps of same sign that occur within 3 hours and interpret them as a single value. The associated residual demand is chosen as the price-weighted average (see Figure 4.5).

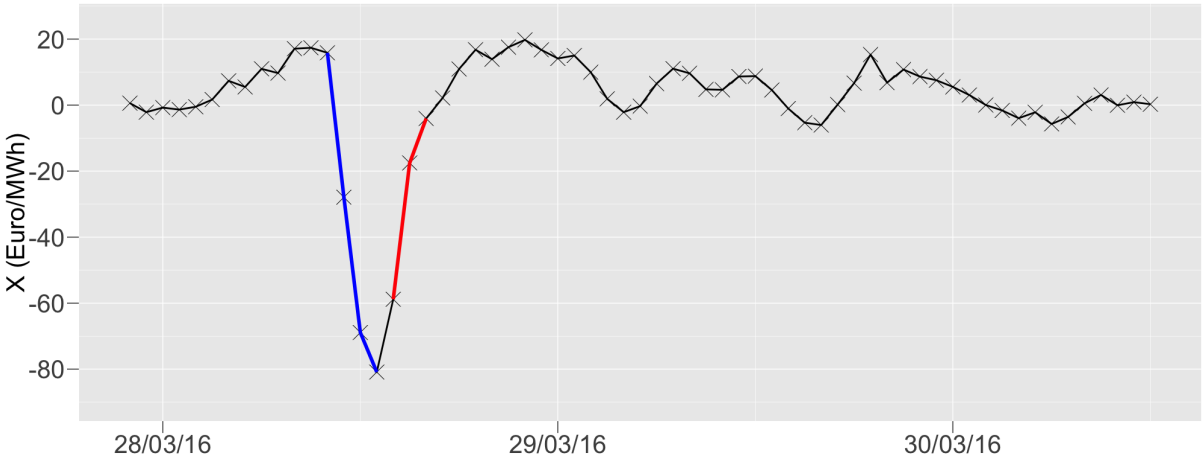

Figure 4.6: Example of detected jumps (blue) and removed jump (red) of the process X (black).

Table 4.5: Testing the null hypothesis that X is a Lévy process.

	m_1	m_2	m_3	m_4
p-value	10^{-16}	10^{-11}	2×10^{-5}	1×10^{-5}

Calibration of the CII process Before we calibrate a R-CII process to the increments $\{\Delta_i X\}_{i=1,\dots,n}$, we want to verify the dependence of X on R. Therefore we apply the test introduced in Subsection 4.1.3 and analyzed in Subsection 4.1.4. The results are stated in Table 4.5. They overwhelmingly point out that a Lévy process is inadequate for modeling X.

To fit the R-CII process X we use the method introduced in Subsection 4.1.2. This requires the specification of an infinitely divisible distribution H_x such that $\Delta_i X \stackrel{d}{=} L(\Delta)$, where $\{L(t)\}_{t \geqslant 0}$ is Lévy process with $L(1) \sim H_x$ for $R(t_{i-1}) = x$ (see Remark 4.4). Assuming that $H_x \approx H_y$ (measured in some proper way) for $x \approx y$, we compare different distributions for the increments $\{\Delta_i X\}_{i \in \mathcal{J}_U}$, where $\mathcal{J}_U = \{2 \leqslant i \leqslant n : R(t_{i-1}) \in U\}$ for some interval U. The length of the interval U is selected in such a way, that the corresponding increments look identically distributed.

Table 4.6 shows the result of Kolmogorow-Smirnow (KS) tests for three common distributions and three intervals, which are chosen according to Figure 4.5. We conclude that the normal-inverse Gaussian (NIG) distribution (see Barndorff-Nielsen (1997)) has enough flexibility to describe the statistical properties of X depending on R. The fact that the NIG distribution fits extraordinarily well on $U = [20, 50]$ confirms that the dependency of X on R is particularly pronounced in case of high or low residual demand.

The NIG distribution allows for tail heaviness and skewness and, therefore, seems to be suitable for modeling a dependence as seen in Figure 4.5. The NIG distribution has parameters α, β, μ and δ, satisfying $0 < |\beta| < \alpha$, $\delta > 0$ and $\mu \in \mathbb{R}$. The parameters μ and δ determine location and scale, respectively, while α and β are influencing tail heaviness and skewness. In particular, the heaviness of the tails is determined by

$$\xi = \left(1 + \delta\sqrt{\alpha^2 - \beta^2}\right)^{-\frac{1}{2}}.$$

The closer ξ is to 1, the heavier are the tails. The parameter

$$\chi = \beta\xi/\alpha,$$

determines the skewness of the distribution. For $\chi < 0$, the left tail is heavier than the right, for $\chi = 0$, the distribution is symmetric, and for $\chi > 0$, the right tail is heavier. Due to the restrictions on the parameters, we always have $0 \leqslant |\chi| < \xi < 1$.

Remark 4.9
In a statistical setting with heavy tails and large skewness, i.e. $|\chi| \approx \xi \approx 1$, the problem of choosing the parameters α and β is ill-conditioned since such a behavior of the density can be obtained for each $\alpha \approx \beta$.

To get estimators for the parameters of the NIG distribution depending on R, we apply the procedure mentioned in Subsection 4.1.2. This requires estimating the first four conditional moments $m_k(x)$, $k = 1, 2, 3, 4$. We use the estimators $\widehat{m}_{k,n}(x)$ as defined in (4.4) with Gaussian kernel, where each bandwidth ϵ_n is chosen according to (4.7). The moments $m_k(x)$ are estimated on the grid $14.7 = x_1 < x_2 < \ldots < x_K = 66.3$ with $|x_i - x_{i-1}| = 0.1$, $2 \leqslant i \leqslant K$, where x_1 resp. x_K correspond to the 1%- resp. 99%-quantile of $\{R(t_i)\}_{i=0,\ldots,n}$.

Figure 4.7 shows the estimated moments along with approximate, asymptotic 95%-confidence intervals (see Theorem 4.6). For positive moments, only those confidence intervals are shown whose lower boundaries are also positive. Especially for the first

Table 4.6: p-value of the KS test.

	Student t	NIG	Normal
$U = [5, 20]$	2×10^{-16}	0.13	2×10^{-16}
$U = [20, 50]$	2×10^{-16}	0.83	2×10^{-16}
$U = [50, 70]$	2×10^{-16}	0.02	2×10^{-16}

moment we see that a constant function does not seem to be suitable for describe $m_1(x)$. This also supports our assumption that X depends on R.

Since the method of moments can be used for fitting of the NIG distribution (see Eriksson et al. (2009)), we transform the estimated moment functions $\widehat{m}_{k,n}(x)$, $k = 1, 2, 3, 4$, into estimates for the parameter functions $\alpha(x)$, $\beta(x)$, $\mu(x)$ and $\delta(x)$. To get parameter functions for all real numbers we linearly interpolate on the grid $\{x_i\}_{1 \leqslant i \leqslant K}$ and extrapolate by the closest estimator. These nonparametric functions are illustrated in Figure 4.8. The resulting skewness function $\chi(x)$ and tail heaviness function $\xi(x)$ are also shown there.

Remark 4.10
The method of moments for $X \sim NIG$ is restricted to the case $3\widehat{\mathcal{K}} > 5\widehat{\mathcal{S}}^2 > 0$, where $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{K}}$ are the sample skewness resp. sample excess kurtosis (see Eriksson et al. (2009)). Estimating the moments $m_k(x)$ on the grid $\{x_i\}_{1 \leqslant i \leqslant K}$, this condition was satisfied for each x_i.

We see that the NIG distribution has heavier tails for low and high residual demand, where the sign of the skewness changes form negative to positive. This is exactly the behavior we already expected by looking at the data (see Figure 4.5).

4.2.4 The long-term variation

The long-term variation Y reflects longer-term changes in the price level, e.g. due to altered commodity prices. Such a change in the price level occurs, for example, between February 2016 and May 2016 where prices increase steadily (see Figure 4.3).

For modeling of Y we follow Burger et al. (2004) and use a *Brownian motion with drift*, i.e.

$$dY(t) = \nu(t)dt + \sigma_Y dW^Y(t), \ t \geqslant 0, \ Y(0) = 0, \tag{4.13}$$

where $\nu : \mathbb{R}^+ \to \mathbb{R}$ is a (possibly stochastic) drift function and $\sigma_Y > 0$ is the volatility of the standard Brownian motion $W^Y = \{W^Y(t)\}_{t \geqslant 0}$. The components are such that $R \perp\!\!\!\perp Y$. Of course, a shift in average residual demand will also change the long-term price level. However, such a change in the level of residual demand is generally seen only in longer periods than those we are considering.

As for Z, we cannot observe Y directly. Of course, for calibration, we could use the moving average filter of Subsection 4.2.3. However, there is also a more sophisticated way of estimating Y. Since $\tilde{S} = S - f(R)$ is a linear function of Y and Z, we can apply

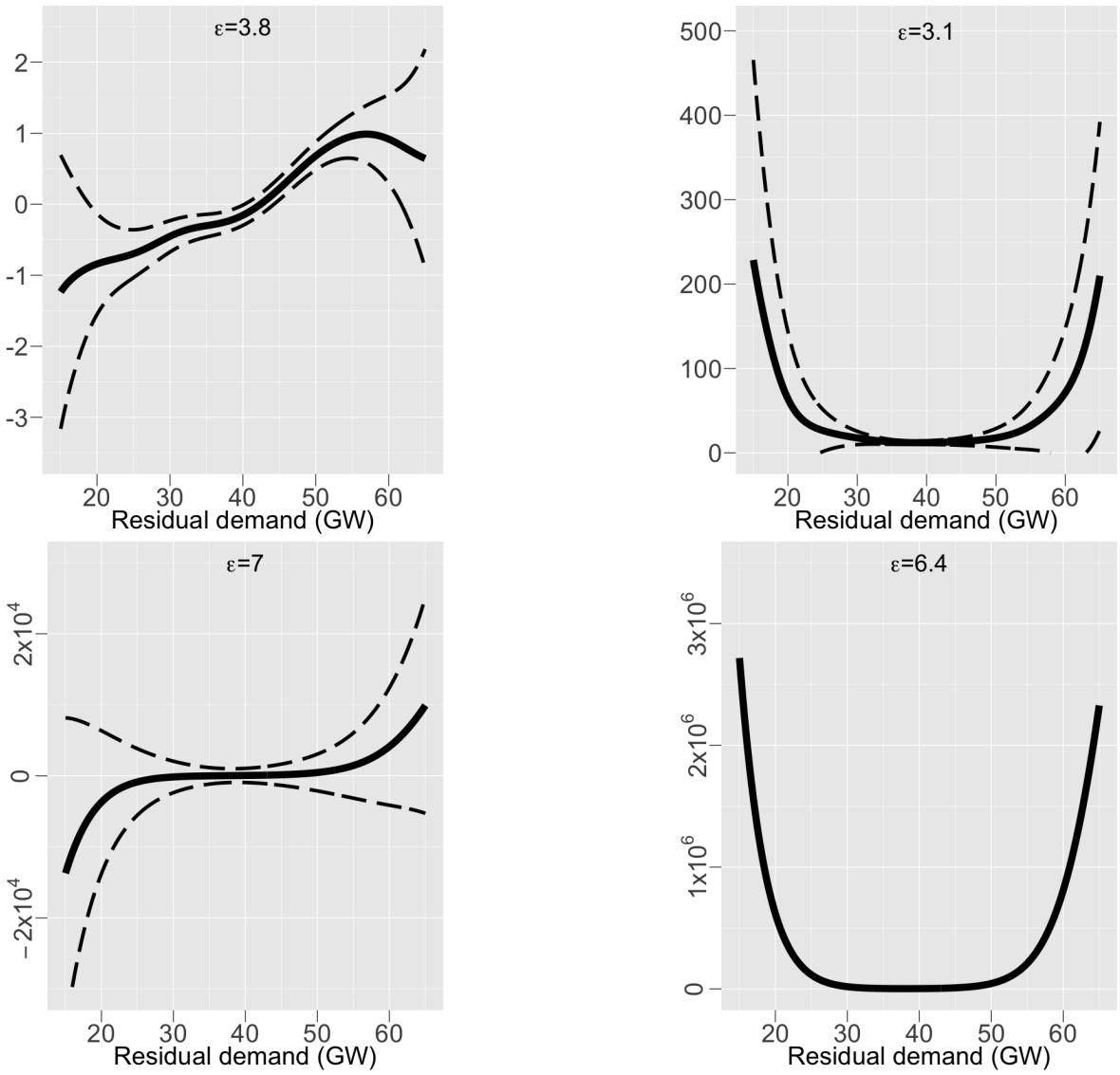

Figure 4.7: Estimated conditional moments $m_k(x)$, $k = 1, 2, 3, 4$, (solid) and 95% confidence intervals (dashed). The chosen bandwidth ϵ is noted in each image.

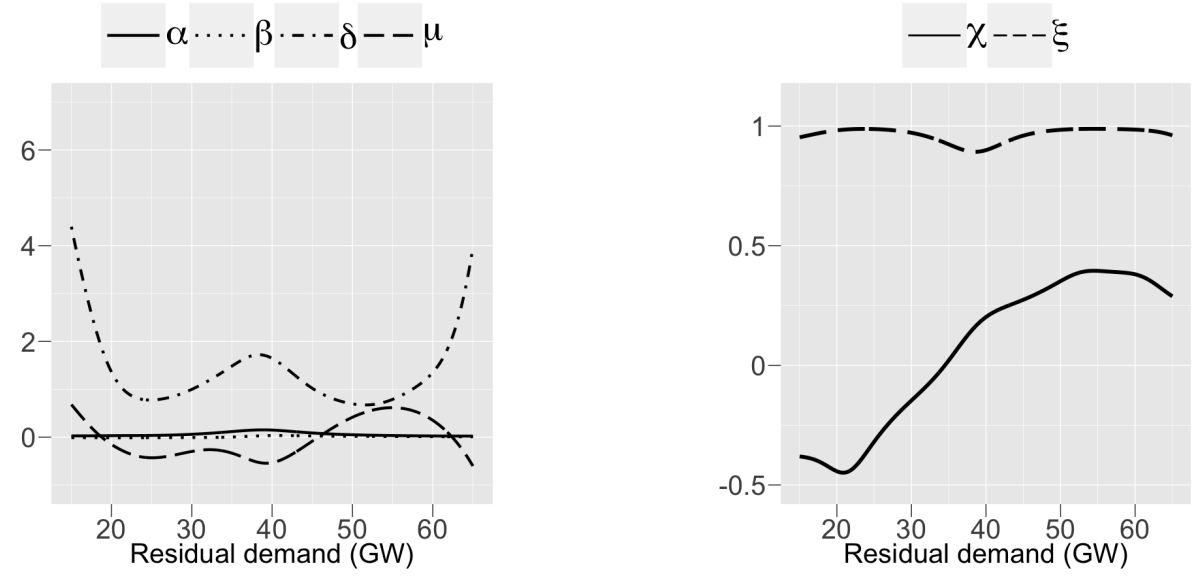

Figure 4.8: (Left) Estimated parameter functions. (Right) Skewness and tail heaviness parameter.

the Kalman filter (see Durbin and Koopman (2012)), to identify Y. This was also done by Burger et al. (2004).

To apply the Kalman filter, we must make an assumption about the drift function ν in (4.13). Since Y represents the long-term price variations, ν should be a smooth function. A simple and easy to handle choice is

$$d\nu(t) = \sigma_\nu dW^\nu(t),\ t \geqslant 0,\ \nu_0 = \bar{\nu}, \tag{4.14}$$

where $\bar{\nu} \in \mathbb{R}$, $\sigma_\nu > 0$ and $W^\nu = \{W^\nu(t)\}_{t \geqslant 0}$ is a standard Brownian Motion independent of W^Y and X. With this Y changes linearly in time and the growth rate may also evolve. Such a dynamic is called a *local linear trend* and is often used in time series analysis. In contrast, Burger et al. (2004) used future contracts with distant maturity to estimate ν. However, such futures are not available for the intraday market.

Discretizing (4.9), (4.13) and (4.14), we employ a state space model

$$\begin{aligned} \tilde{S}(t_i) &= Y(t_i) + Z(t_i),\ i = 1, \ldots, n, \\ \begin{pmatrix} Y(t_i) \\ \nu(t_i) \\ Z(t_i) \end{pmatrix} &= \begin{pmatrix} 1 & \Delta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 - \lambda(Z(t_{i-1}))\Delta \end{pmatrix} \begin{pmatrix} Y(t_{i-1}) \\ \nu(t_{i-1}) \\ Z(t_{i-1}) \end{pmatrix} + \begin{pmatrix} \epsilon_i^Y \\ \epsilon_i^\nu \\ \epsilon_i^Z \end{pmatrix}, \end{aligned} \tag{4.15}$$

where $\epsilon_i^Y \sim \mathcal{N}(0, \sigma_Y^2\Delta)$, $\epsilon_i^\nu \sim \mathcal{N}(0, \sigma_\nu^2\Delta)$ and ϵ_i^Z as in (4.10). All these disturbances are mutually independent. Note that due to its minimal variance property (cf. Durbin and Koopman (2012)[Section 4.2]), the Kalman filter is valid as long as the disturbances are white noise. Hence, the Kalman filter can be applied to derive expected realizations $\{Y(t_i)\}_{i=1,\dots,n}$ based on the observations $\{\tilde{S}(t_i)\}_{i=1,\dots,n}$.

Of course, this requires specification of the unknown model parameters. Since λ and $\text{Var}\left(\epsilon_i^Z \middle| R(t_{i-1})\right)$ are already known by the previous calibrations, we need estimates for σ_Y and σ_ν. These are found by maximum likelihood estimation (see Durbin and Koopman (2012)[Section 7]) and are shown in Table 4.7. The filtered values of Y are shown in Figure 4.3.

Table 4.7: Parameters of the long-term variation Y.

σ_Y	σ_ν
0.11	1.1×10^{-6}

4.2.5 Model analysis

In order to check the statistical quality of our model, we compare model prices with observed prices for the period Jan 01, 2017 to Dec 31, 2017. We restrict the comparison to this year in accordance with our application in Section 4.4. As the main innovation of the model is the inclusion of a CII process to map the short-term price variation accurately, the comparison is made in relation to $S - Y$. In addition, our simulations are based on historical data on residual demand R. This allows to investigate the actual effects of including the CII process driven component Z without distortions caused by an incorrect modeling of R. However, since for application we require realizations of the residual demand, an appropriate model will be given in Section 4.3. The following simulation study is based on 10000 simulated price paths.

Figure 4.9 shows an example of a price path based on the previous calibration. We see that the variability of the ID_3-Price is well-described (cf. Figure 1.9). This is also confirmed by a comparison of summary statistics (Table 4.8). We also noted the moments for the case $Z \equiv 0$. Then, the price depends only on the supply function f, which basically is the model proposed by Wagner (2012). The low value of the standard deviation for the model based on f only shows again, that an additional component is necessary to explain extreme prices.

That especially extreme prices are well described by adding a component driven by a CII

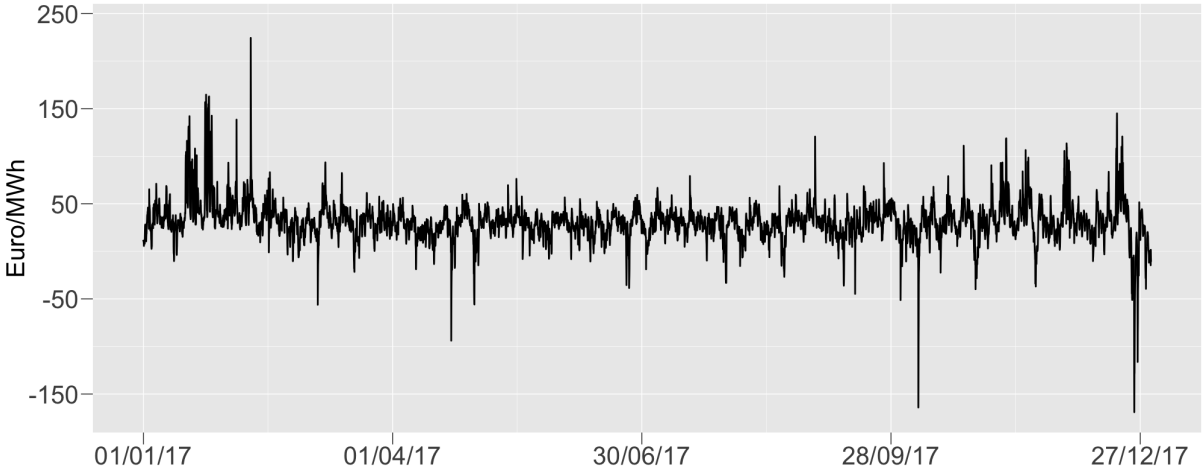

Figure 4.9: Sample path of the price process $S - Y$.

Table 4.8: Summary statistics of $S - Y$.

	mean	st. dev	skewness	kurtosis
Data	31.3	18.9	1	15.8
$f + Z$	31.2	19.8	0.83	10.9
f	31.2	15.0	1.13	9.8

process can also be pointed out by a comparison of the quantiles of the price processes. For this purpose, Table 4.9 opposes empirical upper tail and lower tail quantiles to model quantiles of $S - Y$.

Table 4.9: Comparison of upper tail and lower tail quantiles of $S - Y$.

	0.01%	0.1%	1%	99%	99.9%	99.99%
Data	-119	-84	-11	90	167	216
$f + Z$	-123	-62	-21	98	154	218
f	-52	-38	-10	86	138	150

Figure 4.10 shows the distribution of the simulated prices for the individual hours of Jan 31, 2017. An unusually low production of solar and wind energy led to a price peak of 232 €/MWh on this day. We can see that such an extreme price profile is within the allowable range of our model. Note that, despite the high residual demand, the occurrence of a price peak is by no means predictable. The median of the simulated prices including the short-term variation Z differs insignificantly from the prices found using f only. This reflects the fact that price peaks are unpredictable events, whereas the regular price behavior is completely captured by f.

Figure 4.10 also shows that in addition to the intraday seasonality with low prices at night and higher prices during the day, the variance of the price also changes accordingly. Hence, by modeling Z driven by a R-CII process, a time-varying volatility is described, which must be modeled separately for more classical stochastic processes.

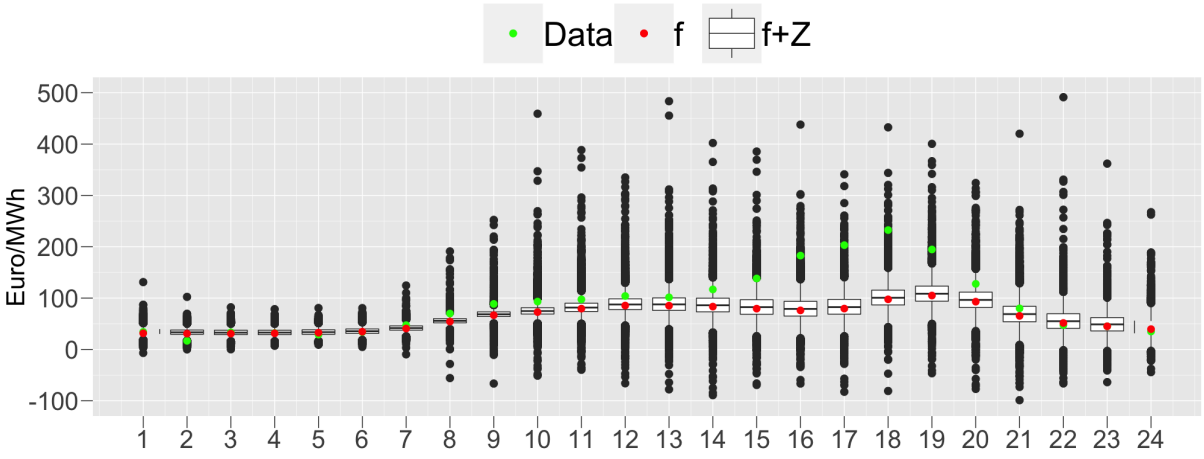

Figure 4.10: Boxplot of the simulated price $S - Y$ for the day of Jan 31, 2017.

4.3 A residual demand model

So far we have not discussed a model for the residual demand R, although this is obviously essential for the calculation of the electricity price S in our model. To characterize R we need a model for electricity demand and infeed from renewables. Since we focused on solar and wind power, it is enough to consider these two renewable energy sources, i.e.

$$R(t) = D(t) - S(t) - W(t), \quad t \geqslant 0, \tag{4.16}$$

where $D = \{D(t)\}_{t\geqslant 0}$ denotes total load, $S = \{S(t)\}_{t\geqslant 0}$ denotes solar power infeed and $W = \{W(t)\}_{t\geqslant 0}$ denotes wind power infeed. Figure 4.11 shows historical residual demand in Germany for the years 2016/2017.

Total load Electricity demand has a pronounced seasonality on different time scales (cf. Section 1.3). For D we therefore assume an arithmetic model consisting of a deterministic trend and seasonality function Λ^D and a noise process $X^D = \{X^D(t)\}_{t\geqslant 0}$,

$$D(t) = \Lambda^D(t) + X^D(t), \quad t \geqslant 0. \tag{4.17}$$

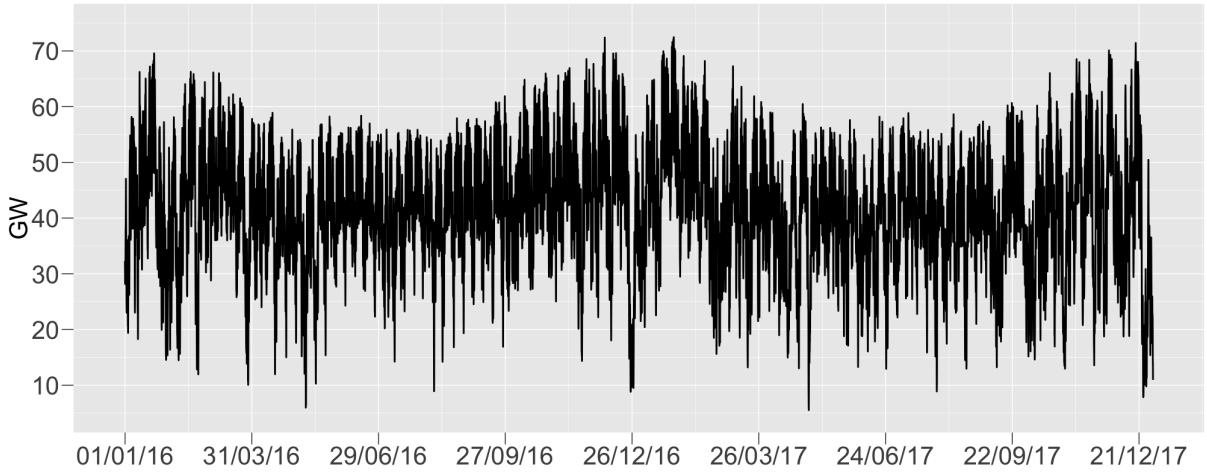

Figure 4.11: Residual demand as total demand reduced by infeed from solar and wind power for the years 2016/2017.

First, to depict trend and yearly seasonality of D, we use a linear term and a sum of trigonometric functions

$$y(t) = c_0 + d_0 t + \sum_{k=1}^{K} c_k \cos\left(\frac{k2\pi t}{365 \times 24} + d_k\right), \tag{4.18}$$

where $1 \leqslant K \leqslant 182$ is the number of harmonics and time t is measured in hours. We fit this function for different K to the average demand on business days and found that only the first three sinusoidal waves are significant. Figure 4.12 shows that $K = 3$ gives a parsimonious representation of the yearly seasonality. The parameter estimates of this function are shown in Table 4.10.

Table 4.10: Least squares estimates of the parameters of the trend and seasonality function Λ^D.

c_0	d_0	c_1	d_1	c_2	d_2	c_3	d_3	w_1	w_2
58	2×10^{-4}	3.9	-0.2	1.2	0.2	-0.8	0.2	14	9

To map the remaining seasonality we follow an approach similar to that of Wagner (2012). Lower demand on weekend and holiday is accounted for by constant factors w_1 and w_2,

$$\psi(t) = \left(1 - w_1 \mathbf{1}_{\{w_1 \in B_1\}} - w_2 \mathbf{1}_{\{w_2 \in B_2\}}\right) y(t),$$

where B_1 is the set of all public holidays and Sundays and and B_2 is the set of all partial holidays and Saturdays. w_1 and w_2 are calculated as the mean deviation of D from y (see Table 4.10). Finally, the intraday seasonality is reflected by daily pattern

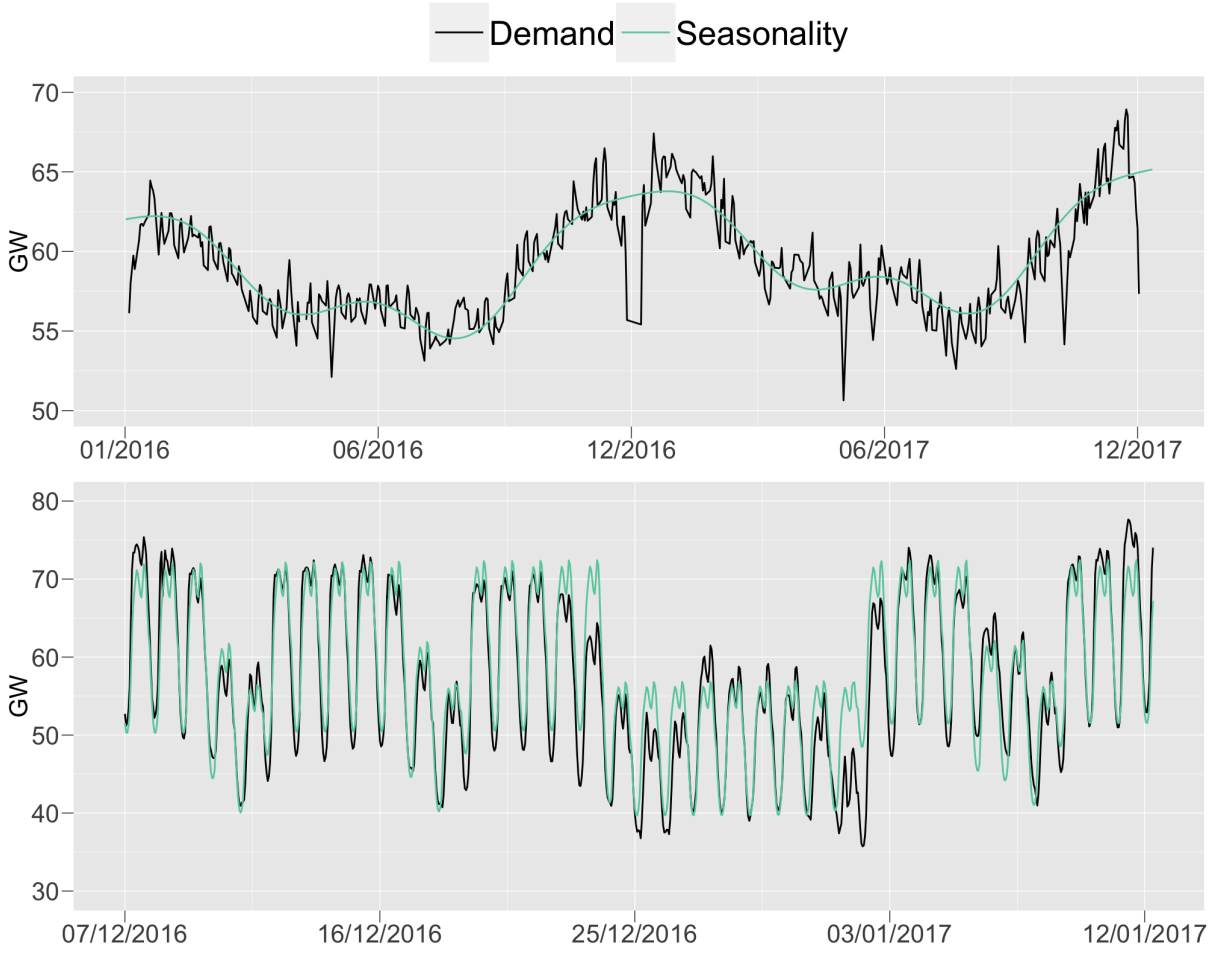

Figure 4.12: (Upper) Average demand on business days and fitted trend and yearly seasonality function y. (Lower) Demand and fitted trend and seasonality function Λ^D.

transformations δ_m, $m = 1, 2, \ldots, 12$, similar to those in (1.3), where we set the coefficients $c_{m,k}$, $k = 1, 2, \ldots, 24$, as the mean of the ratio between the actual load and the daily average load.

Therefore, electricity demand has a trend and seasonality function of the form

$$\Lambda^D(t) = \sum_{m=1}^{12} \delta_m(t, \psi(t)) \mathbf{1}_{\{t \in (\text{m-th month of the year})\}}.$$

Figure 4.12 illustrates that this function mirrors the seasonality of electricity demand on different time scales quite well.

By subtracting the estimated trend and seasonality function, we obtain the noise process X^D. Since electricity demand is strongly connected to mean-reverting factors like temperature, we expect that X^D also has the mean-reverting property. This assump-

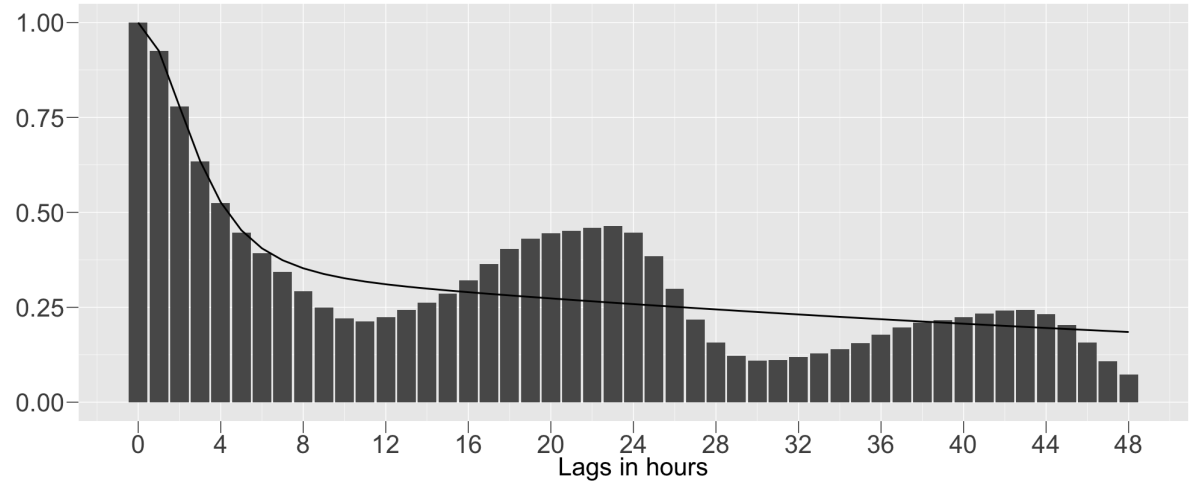

Figure 4.13: Empirical ACF of the noise process X^D (bar) and ACF of a fitted CARMA(5,4) process (solid).

tion is supported by an investigation of the empirical autocorrelation function (ACF) (cf. Figure 4.13).

We also observe that the empirical ACF is not a monotone function. In addition to a strong decay at the beginning, there seems to be a significant dependency between two realizations with a lag of about one day. This could be attributed to fluctuations in demand due to actual weather conditions. Since for many weather-related phenomena a daily seasonality can be observed, this translates to the autoregressive structure of demand. We also see that the third peak of the empirical ACF is not exactly at twice the lag of the second peak. This also indicates that the observed dependence is not part of a daily deterministic seasonality. To model such complex dependency structures, continuous-time autoregressive moving average (CARMA) processes are well suited.

In Section 1.5 we discussed the CARMA(p, q) process of order $p, q \in \mathbb{N}$, $q < p$, in detail and also dealt with its estimation and simulation. What we have not addressed so far is the choice of an appropriate model order, i.e. the selection of p and q. Tómasson (2015) recommends to only consider CARMA$(p, p-1)$ processes for different values of p. This has the advantage, that nested models are produced and usual model selection criteria can be used. Following this advice, Table 4.11 states the maximized Gaussian log-likelihood of different CARMA$(p, p-1)$ processes for X^D.

We conclude that a CARMA$(5, 4)$ process is appropriate, applying e.g. the Bayesian information criterion (BIC). After identifying the CARMA parameters by quasi maximum likelihood estimation, we recovered the increments of the driving process (cf. Subsection 1.5.2). There were some significant autocorrelation for this increments especially

Table 4.11: Maximized Gaussian log-likelihood of CARMA$(p, p-1)$ models.

	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=7$
X^D	-23852	-20407	-20305	-20175	-20156	-20154	-20151
X^W	10306	18148	18263	18265	18266	18266	18267

for a lag of one day. However, we will assume that the driving process has independent increments. Following the result of a KS test (see Table 4.12), we conclude that we have a Lévy process L^D of NIG type.

Table 4.12: p-value of the KS test for recovered increments of the driving Lévy processes.

	Student t	NIG	Normal
L^D	2×10^{-16}	0.79	2×10^{-16}
L^W	2×10^{-16}	0.72	2×10^{-16}

The parameter estimates are given in Table 4.13. As it is common in numerical application, we fixed b_0 to the value of 1 (cf. Remark 1.1). The parameters of the NIG process L^D (normalized to time 1) are noted as α^D, β^D, δ^D and μ^D. The associated NIG distribution has a tail heaviness of $\xi^D = 0.69$ and a skewness of $\chi^D = -0.02$ indicating slightly heavier tails than for a normal distribution and symmetry.

Table 4.13: Parameters of the NIG CARMA$(5,4)$ process X^D.

a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	α^D	β^D	δ^D	μ^D
23.8	50.8	44.4	12.1	0.2	20.3	9.4	1.9	0.5	0.35	-0.01	3.14	0.1

In Figure 4.13 we have drawn the ACF of the fitted CARMA process. Obviously, different mean-reversion rates are considered, but the daily seasonality is not clearly expressed. However, here it must be taken into account that parameter estimates are found by the maximum likelihood method and not by fitting simply the ACF.

Figure 4.14 shows a sample path of D as the sum of the trend and seasonality function Λ^D and the fitted noise process X^D. The behavior of D (cf. Figure 1.1 and Figure 4.12) seems to be properly reproduced.

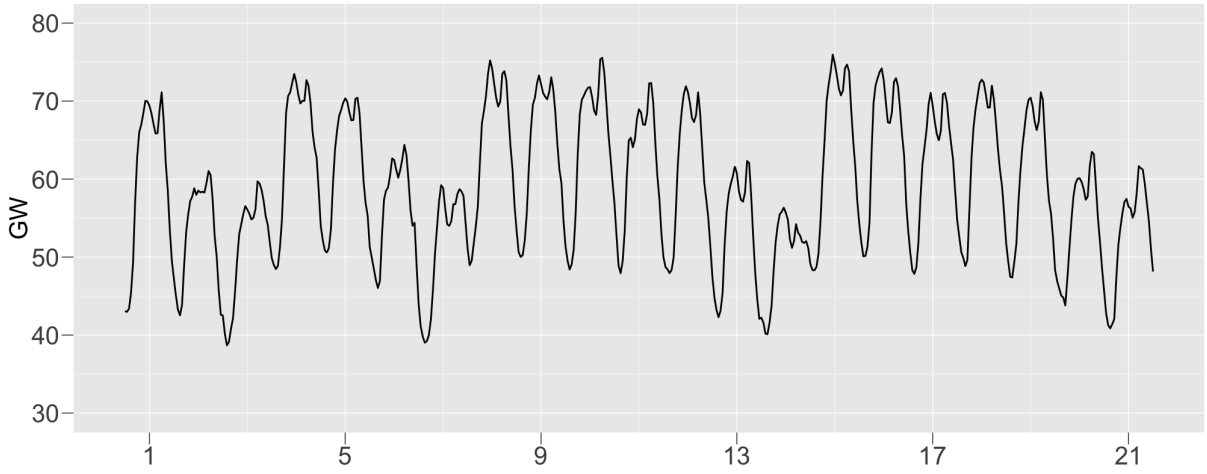

Figure 4.14: Simulation of the total load D for three consecutive weeks covering the period Oct. 30, 2017 to Nov. 19, 2017.

Table 4.14: Least squares estimates of the parameters of the seasonality function Λ^W.

c_0	c_1	d_1	c_2	d_2	c_3	d_3
-1.6	0.5	-19	0.1	1	-0.1	5.5

Wind power infeed Stochastic models for wind power have already been briefly mentioned for structural electricity pricing in Section 1.4. Both in the residual demand model of Wagner (2012) and for the Cox process utilized by Deschatre and Veraart (2017) an arithmetic model was applied to a logit-transformed wind power variable. For the noise term, Gaussian continuous-time autoregressive (CAR) processes were used. In contrast, Benth and Pircalabu (2018) used an Ornstein-Uhlenbeck (OU) process driven by a subordinator to ensure that an lower bound is never exceeded. They considered the negative logarithmized time series of wind power efficiency (cf. Equation 1.2), which is bounded by 0.

We will follow the direct approach and fit an arithmetic model as in (4.17) for the logit-transformed wind power efficiency. Wind power has a strong annual pattern (cf. Figure 1.5). Since no change of the efficiency of wind power plants is to be expected in the medium term, we use a sum of trigonometric functions as in (4.18) with a drift of zero as seasonality function Λ^W. A number of $K = 3$ sinusoidal waves was found to be appropriate. Figure 4.15 shows the fitted function. Their parameters are stated in Table 4.14.
For the noise process X^W, the same procedure as for electricity demand is applied. First, a CARMA(3,2) process is identified as proper model by comparing the log-likelihood values (cf. Table 4.11). The estimated parameters are shown in Table 4.15, where the recovered increments again indicate an NIG Lévy process as driving noise (cf. Table 4.12).

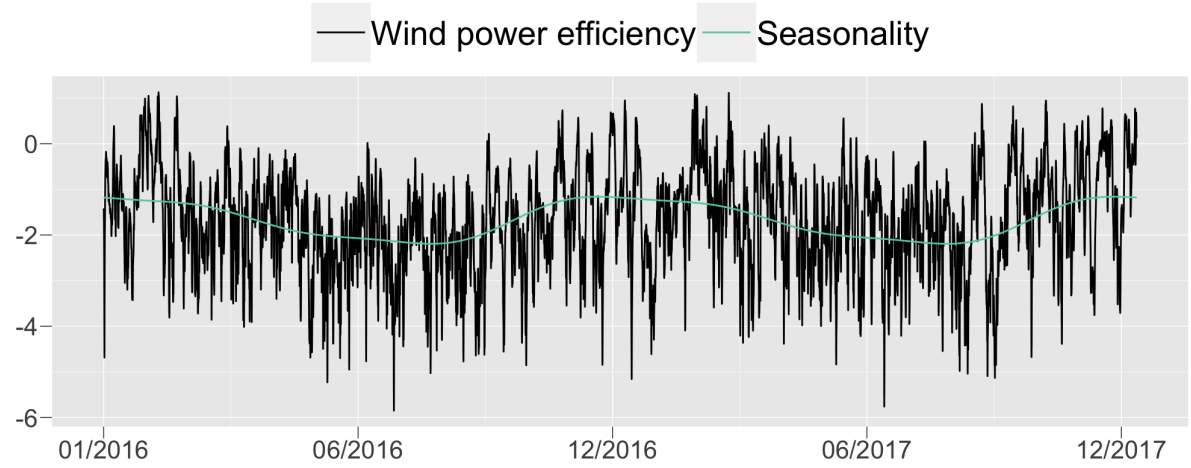

Figure 4.15: Logit-transformed wind power efficiency and fitted seasonality function Λ^W.

It is not surprising that the distribution of the NIG Lévy process L^W has heavier tails ($\xi^W = 0.79$) than we got for the driving noise of the demand process. The skewness is again negligible ($\chi^W = 0.07$).

Table 4.15: Parameters of the NIG CARMA$(3,2)$ process X^W.

a_1	a_2	a_3	b_1	b_2	α^W	β^W	δ^W	μ^W
4.2	2.3	0.08	0.22	0.09	1.28	0.11	0.49	-0.04

Figure 4.16 shows that by a CARMA(3,2) process the autoregressive structure of X^W is nicely reproduced. A sample path of W is given in Figure 4.17 which appears to capture the characteristics of wind power infeed (cf. Figure 1.6).

Residual demand To fully characterize the residual demand R, it remains to define a model for the solar power infeed S. Such a model has already been discussed in detail in Chapter 3. If we combine this with the previously defined models for the total load D and the wind power infeed W, by (4.16), we get the desired process R.

In Figure 4.18 a sample path of the compound residual demand process for the years 2016/2017 is illustrated. The simulated values are in agreement with the true process (cf. Figure 4.11). Note that there are two days with an oversupply of renewable energy. Although we did not observe this for our data, such a situation cannot be ruled out in general. The surplus of electricity would be transported to other European countries in that case.

A comparison of summary statistics (cf. Table 4.16) also indicates that our residual demand model is in line with the data, where a simulation study based on 125 simulated

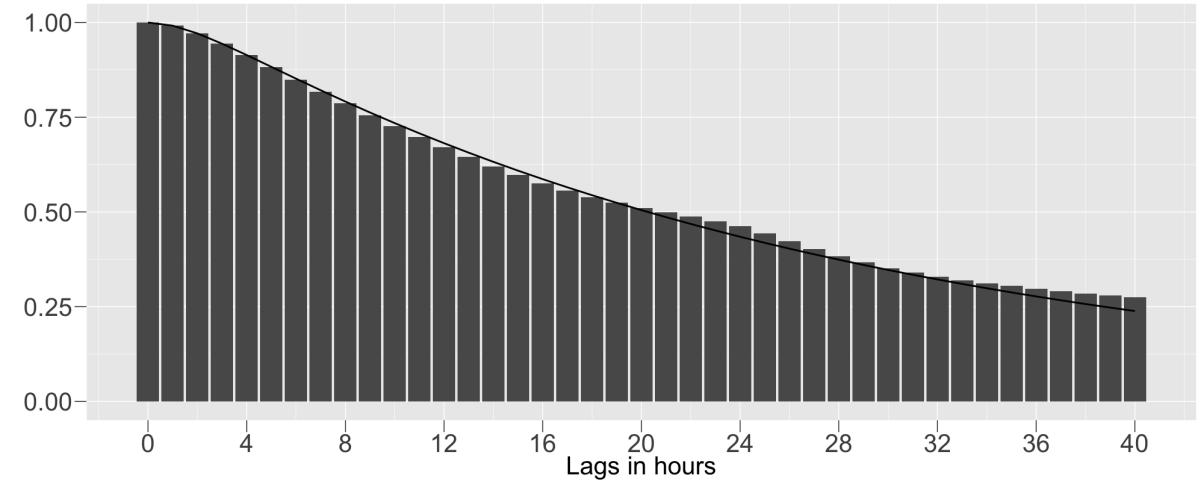

Figure 4.16: Empirical ACF of the noise process X^W (bar) and ACF of a fitted CARMA(3,2) process (solid).

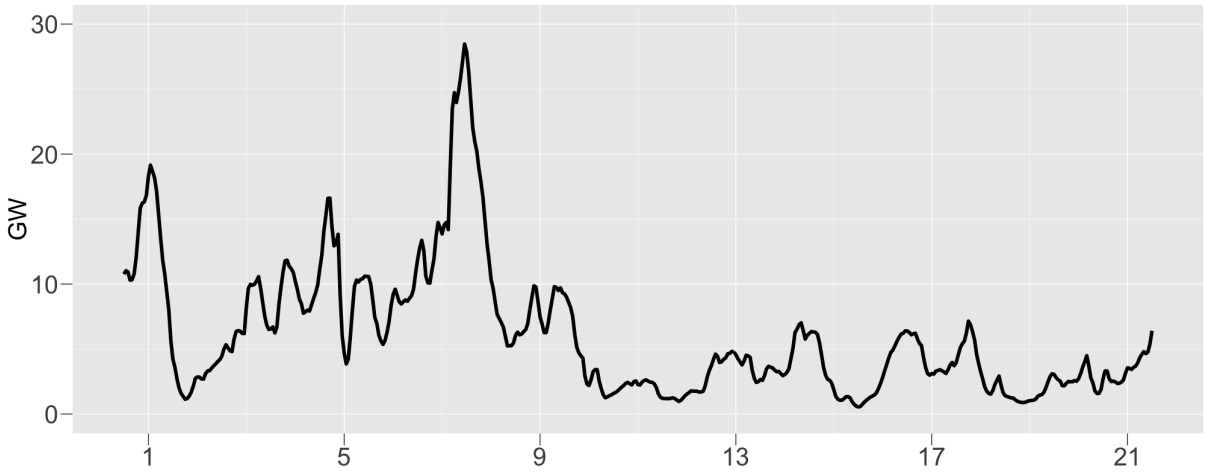

Figure 4.17: Sample path of wind power infeed W for three consecutive weeks covering the period Apr. 30, 2017 to May 20, 2017.

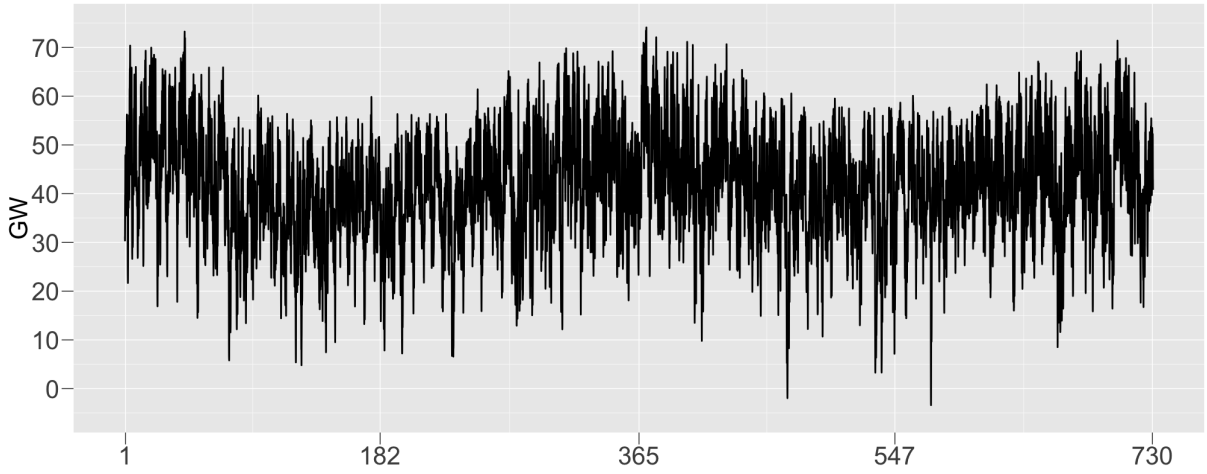

Figure 4.18: Sample path of the compound residual demand $R = D - S - W$ for the years 2016/2017.

Table 4.16: Summary statistics of the residual demand R.

	mean	st. dev	skewness	kurtosis
Data	41.28	11.32	-0.09	2.77
Model	41.34	11.41	-0.31	3.19

path for the years 2016/2017 is applied to approximate the moments of the residual demand model. The empirical standard deviation is about 20% for the skewness and less than 5% for the other statistics.

The proposed residual demand model allows a simple simulation of the process R. This is helpful for application of our electricity price model defined in Section 4.2. Here it must be noted, that the residual demand model of this section is not of the form we proposed for the external factor of a CII process (cf. Equation 4.3). However, this form was only required to ensure that the external factor spends enough time at the points we want to estimate the conditional moments of $\Delta_i X$. But, by the mean reversion and strong seasonality of R, we have no doubt that this also applies here.

4.4 Valuation of power cap and floor

High or low infeed of renewable energies can lead to extreme market prices. Therefore, by the expansion of renewable sources, all operators of power plants are exposed to new type of price risks. In order to provide an opportunity to hedge against these risks, the EEX recently introduced products based on the ID_3-Price called *German Intraday Cap/Floor Futures*.

Based on the models of Section 4.2 and Section 4.3, we now want to analyze prices of these derivatives. We use data for the period Oct 03, 2016 to Jan 28, 2018 provided by the EEX. Data are only available for the floor from Mar 15, 2017. The published prices are the daily settlement values for Monday to Friday.

Table 4.17: Contract specification of the German Intraday Cap/Floor Futures.

Underlying	EPEX SPOT Intraday index (ID_3-Price)
Strike	40 € per MWh (Cap) and 10 € per MWh (Floor)
Contract size	168 MWh
Maturities	Week contracts (current week + 4 weeks ahead)
Market area	Germany
Fulfillment	Cash Settlement

A cap is a derivative where by buying at time s a payoff of the difference between the market price $S(t)$ and the strike price K over the delivery period $t \in [\tau_1, \tau_2]$ is received, i.e. a buyer receives

$$\sum_{t=\tau_1}^{\tau_2} \max\left(S(t) - K, 0\right) - (\tau_2 - \tau_1 + 1) F^{\text{cap}}_{\tau_1,\tau_2}(s), \tag{4.19}$$

where $F^{\text{cap}}_{s\tau_1,\tau_2}$ is the cap price at time s. The payoff from being long a floor is calculated accordingly as

$$\sum_{t=\tau_1}^{\tau_2} \max\left(K - S(t), 0\right) - (\tau_2 - \tau_1 + 1) F^{\text{floor}}_{\tau_1,\tau_2}(s). \tag{4.20}$$

The specifications for the cap and floor traded at the EEX are shown in Table 4.17. The chosen strikes correspond to the 85% quantile resp. 5% quantile of the ID_3-Price from 2015. Hence, the German Intraday Cap/Floor Futures allow to hedge the risk of high or low prices on the intraday market.

In contrast to classical mathematical finance, in which derivatives are priced based on risk-neutral arguments, since buy-and-hold stategies are not applicable, in the electricity market valuation is usually based on rational expectations (Benth et al. (2008)[Chapter 1.5.2]). This means that traders base their bids on the expected price. Hence, for a cap with payment as in (4.19), we have

$$\mathbb{E}\left(\sum_{t=\tau_1}^{\tau_2} \max\left(S(t) - K, 0\right) \middle| \mathcal{F}_s\right) - (\tau_2 - \tau_1 + 1)\left(F^{\text{cap}}_{\tau_1,\tau_2}(s) - RP_{\tau_1,\tau_2}(s)\right) = 0,$$

where $\{\mathcal{F}_s\}_{s\geqslant 0}$ is the filtration of all market information up to time s and RP_{τ_1,τ_2} is a risk premium. This leads to a cap price

$$F^{\text{cap}}_{\tau_1,\tau_2}(s) = \frac{1}{\tau_2-\tau_1+1}\mathbb{E}\left(\sum_{t=\tau_1}^{\tau_2}\max\left(S(t)-K,0\right)\middle|\mathcal{F}_s\right) + RP_{\tau_1,\tau_2}(s).$$

The price of the floor is found accordingly using (4.20).

Figure 4.19 shows the seasonal behavior of the German Intraday Cap/Floor Future prices. Since we assume that no relevant information about the occurrence of price peaks will be

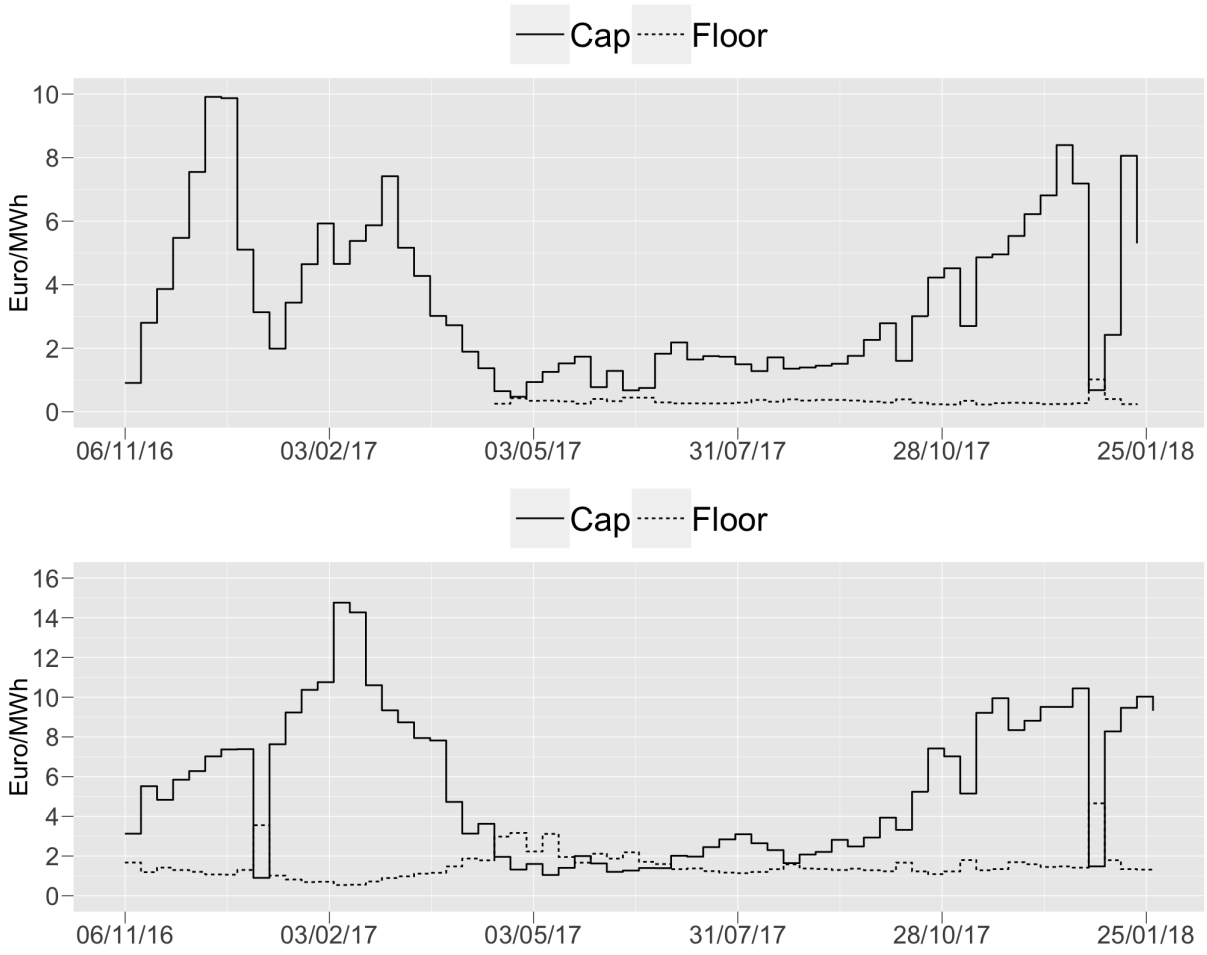

Figure 4.19: Observed (upper) resp. simulated (lower) German Intraday Cap / Floor Future prices with five weeks to maturity.

available already four weeks in advance, we can use the first quoted price of each product to determine the annual pattern of the price. For the cap, we see that positive price peaks are more likely to be expected in winter. This can be explained by the higher demand in Germany during the cold season (see Figure 1.1). An exception is the weeks around

Christmas, which represent a main holiday season. We notice that here the expected payoffs are dropping to a very low level. For the floor, the situation is mirrored, with prices being on average only about one tenth of the cap's prices. This is not surprising, as the chosen strikes lead to more positive payments of the caps than of the floors. This is further enhanced by the fact that in 2017 there is generally a higher price level than before (see Figure 4.3).

Using the residual demand process of Section 4.3, we simulated 10000 realizations of the model as proposed in Section 4.2 for each week of the observation period starting five weeks before settlement. We used this for a Monte Carlo approximation of the expected derivative payments (see Figure 4.19). The sample standard deviation was not larger than 6. The model prices show a similar seasonal behavior as the observed prices. Also the strong change of the prices around Christmas is to be recognized.

However, there are two differences. First, the observed cap prices until February 2017 show a bimodal behavior in contrast to the simulated prices. This can be traced back to an unusual non-stationary variation of the ID_3-Price that is not attributable to residual demand and is not captured by the long term variation (see Figure 4.3). Second, we can see that the expected payments are greater than the derivative prices.

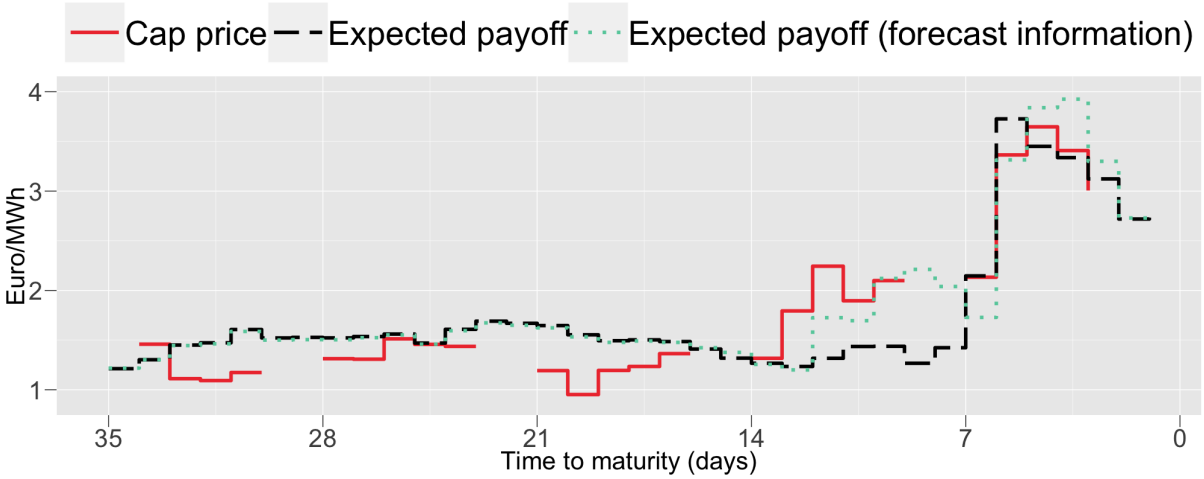

Figure 4.20: Price of the cap and expected payment for the delivery period May 29, 2017 to Jun 04, 2017 depending on the time until maturity.

The fact that the expected payments are larger than the future prices indicates a negative risk premium. In order to examine the risk premium more closely, we use a Monte Carlo study as before. We calculate the expected payment for each week of delivery depending on the time to maturity. Because of the unusual behavior of the ID_3-Price before March 2017, we use the products from Mar 05, 2017 for our study.

As an example, Figure 4.20 shows the price evolution of the cap with delivery period as the 22 calendar week 2017. By averaging across all differences between the derivative prices and the expected payoffs, we get an idea how the risk premium behaves.

The results are shown in Table 4.18. Our previous hypothesis of a negative risk

Table 4.18: Mean risk premium RP_{τ_1,τ_2} by weeks to maturity.

	Five	Four	Three	Two	One
Cap	-0.91	-0.75	-0.66	-0.75	-0.21
Floor	-1.56	-1.56	-1.56	-1.42	-0.89

premium is generally confirmed, implying that the market is in backwardation. The fact that a discount is given on the expected payoff for both the hedge against high and low prices is surprising at first. A possible explanation could be that conventional producers, such as gas-fired power plants, adjust their production to prices on the intraday market. In the case of high prices they increase production to generate additional income, while in the case of low prices they reduce production and fulfill their obligations by using electricity bought on the market. To hedge their revenues they sell caps and floors.

On the other hand, renewable power generators face the risk of over- or under-production compared to day-ahead market commitments. Since it can be assumed that a substantial overproduction often correlates with a high overall production of renewable energies and this leads to low prices and vice versa, renewable power generators must buy caps and floors for hedging purposes. However, as commitments for the sale of renewable energy are made only for short periods in advance, the purchase of the derivatives takes place closer to maturity. This corresponds to the increasing risk premia as we approach maturity.

To find the seasonality of the derivative prices, market information does not play an important role. However, for times close to the delivery period, market participants have access to information about the occurrence of extreme prices. Forecasts on wind and solar power generation seem particularly important here. This assumption is also supported by the unexpected change of the cap price in Figure 4.20 one week ahead the delivery period.

The incorporation of such information in our model is straightforward. Instead of simulating the price based on the realization of the residual demand model, we simulate using the realization of a forecast distribution. The simulated derivative prices, using such an information, seem to better correspond to the observed prices (see Figure 4.20). Here, we let depend the residual demand on hourly predictions of the total amount of wind

and solar power generated in Germany for the first six days from the time of simulation, after which we again use the residual demand model of Section 4.3. The forecast data was provided to us by EuroWind GmbH, an accredited provider of prediction data that also cooperates with EEX.

However, if such information is used by market participants this means that the perviously calculated risk premia are not reliable for days close to maturity. The need to include forecast information was also addressed in Benth and Pircalabu (2018) for the computation of the risk premium for wind power futures.

Table 4.19: Mean risk premium RP_{τ_1,τ_2} by weeks to maturity using forecast information.

	Five	Four	Three	Two	One
Cap	-0.91	-0.75	-0.67	-0.71	-0.07
Floor	-1.49	-1.49	-1.48	-1.19	-0.40

Thus, in Table 4.19 the risk premia using the given predictions are shown. We can see that the previous statements on the risk premium still apply. However, the increase as it approaches maturity is now more pronounced. This supports the earlier thesis that for these times more producers of renewable energies will appear as buyers in order to hedge their obligations from recent commitments.

Chapter 5

Summary and Conclusion

The intention of this thesis was to develop accurate stochastic models for electricity markets. Our focus was on continuous time autoregressive processes that have been applied successfully many times in this field. Even though our motivation was data driven and modeling was the final goal, the methods introduced were always subject to a solid mathematical analysis.

The mean-reversion within short periods of time is probably the most distinctive characteristic of electricity prices. In particular, jumps only have a very short impact on the price. We have therefore started our research with a more detailed investigation of this property. It is generally known that the mean-reversion rate of electricity prices depends on the price level. Hence, in Benth et al. (2008) an arithmetic price model consisting of two factors with different mean-reversion rates is proposed. The main difficulty here is to separate the price dynamic into a normal and a jump part. As attractive alternative Garcia et al. (2011) identified continuous-time autoregressive moving average (CARMA) processes. Benth et al. (2014) stated that by using a higher order of the autoregressive part, different mean-reversion speeds can be included here. However, since the price level has no direct impact on the mean-reversion rates, we concluded that there is still some potential for improvement.

To get mean-reversion rates depending on the price level, it appears to be natural to consider parameter functions instead of constants. For CARMA process such an idea leads to the class of continuous-time threshold autoregressions (CTAR). Since Brockwell (1994) only considered Gaussian CTAR, in Chapter 2 we extended this concept by adding a jump component. This is necessary since jumps are frequently observed for electricity prices. In addition to proving the existence of the newly defined process, we have introduced a consistent Euler method which can be used for simulation. While these points represent extensions of the ideas used by Brockwell (1994), we have also analyzed an estimation

procedure that has not yet been considered for CTAR. The method utilizes particle filtering in case of analytically intractable densities. By a simulation study we verified the quality of the estimation procedure. Finally, we have shown that CTAR with jumps can offer a significant advantage over comparable models in describing electricity spot prices.

Since CTAR are purely autoregressive, it would certainly be interesting to add also a moving average part. However, as shown by Brockwell (1994), this leads to a multivariate stochastic differential equation (SDE) with degenerate diffusion matrix. This case is much more complicated than for CTAR and results cannot simply be extended. Also the general statistical properties of CTAR could be investigated further, especially for a CTAR with jumps. This includes not at least a convergence analysis of the likelihood estimate for discontinuous transition densities (cf. Remark 2.15).

CTAR are not only suitable for describing electricity prices, but have also proved useful in modeling photovoltaic (PV) power generation. Solar power is one of the most important renewable energy sources and is recognized as a key technology to fight climate change. However, due to its variable generation profile, it also presents the electricity market with new challenges. Exact models for solar power infeed are essential to cope with them. Since the amount of generated solar power depends heavily on the time of day, for classical statistical modeling its time series has to be stationarized first. This is difficult because infeed is zero at night. Therefore previous models have been limited to daily data (cf. Wagner (2012), Veraart and Zdanowicz (2016) and Benth and Ibrahim (2017)).

In Chapter 3 we proposed a statistical model for PV power generation, where, for the first time, intraday variation is taken into account. This was possible, because instead of considering the time series directly, we identified the cloud cover as essential component. We first removed seasonality by estimating the clear sky infeed. Then, the remaining variation can be interpreted as a cloud cover component, a very natural view of the stochasticity of solar power generation. The filtered cloud cover showed to be still non-stationary and highly asymmetric. To achieve stationarity, we first transformed the time series by using physical relationships. To model the asymmetry, CTAR were identified as appropriate class of processes. Using a continuous time model has the additional benefit here, that missing values can be easily handled. This is necessary because it is not possible to get information on the cloud cover at night by infeed data only. We fitted the process with the procedure presented in Chapter 2. The model subsequently turned out to be superior to an existing concept. The purely statistical approach makes it easy to use, e.g. for financial application. To illustrate this, we introduced a new future contract that can be used to hedge volume risks from PV power plants.

Although there is no direct price risk for power plants subsidized under the Gesetz für den Ausbau erneuerbarer Energien (Renewable Energy Sources Act) (EEG), additional revenues can be generated through efficient commercialization. A very efficient variant is the storage of generated electricity in low price periods and the sale at high prices. To control such a strategy, the costs of the storage must be compared to the potential yields. Intraday generation of the power plant plays an import role here. Hence, it would also be interesting to use the proposed model for the valuation of a battery storage.

Precise models for renewable power generation are not only important for operators of associated power plants, but for many other market participants as well. The production of electricity from renewable energy sources has vanishing marginal cost. In a liberalized market, they therefore replace conventional power plants and have a strong impact on the price. Recent price models (cf. Section 1.4) attempt to take this impact into account. The main focus here is on applying existing stochastic techniques, which, however, usually only allow to map certain aspects of the dependency between the price and renewable power generation.

Conditionally independent increment (CII) processes provide a flexible way to represent many parts of this dependency. While this class of processes has generally been known for a long time (cf. Jacod and Shiryaev (2002)[II.6]), it has not yet been investigated for modeling of time series. This is discussed in Chapter 4. Besides dealing with the construction of a CII process, we also provide an estimation procedure. In addition, we built a test to check if a simpler Lévy process might not be already sufficient. The estimates and the test showed a good performance in a simulation study.

Based on the new theoretical results about the statistical application of CII processes, we defined an arithmetic electricity price model for the German intraday market. We first mapped residual demand to the price by a deterministic function. With this much of the seasonality and regular variation of the price is already reflected. To take into account changes in the long-term price level, a Brownian motion with stochastic drift as additional factor is proposed. Filtering and estimation of this factor follows from the theory of linear state-space model. By the components of the model we mentioned so far, especially price peaks are not sufficiently represented. However, these are of crucial importance for the assessment of price risks. Since jumps only have a short effect on the price, a generalized Ornstein-Uhlenbeck (OU) process is suggested for the variations not covered so far. For the increments of this process we found that not only the sign but also the magnitude can be explained in terms of residual demand. This was to be expected as price jumps often are the result of an extraordinarily high or low renewable power generation. By applying a CII process such a dependency can be depicted. After

some data preprocessing, a normal-inverse Gaussian (NIG) distribution with parameters depending on residual demand was fitted using the previously established nonparametric estimator. A simulation study illustrated that by the inclusion of a CII process more reasonable prices are achieved, especially if one is interested in extreme values, compared to other models.

By our approach, electricity prices are connected to load as well as solar and wind power generation. Hence, for a complete electricity price model these components must also be specified. Modeling of solar power generation was already treated in detail in Chapter 3. For electricity demand and wind power generation arithmetic models based on Lévy driven CARMA processes were applied (cf. Section 4.3).

The EEX recently offers the German Intraday Cap/Floor Futures for hedging against positive and negative spikes on the intraday market. The fact that our proposed electricity price model seems to replicate such spikes quite well, gave reason to analyze the prices of this derivatives in more detail. We have found signs of a negative risk premium with a decreasing magnitude if maturity is closer. An explanation could be that conventional power plants in particular are hedging their revenues on the intraday market. For calculation of the risk premia we also took forecast information on wind and solar power generation into account, which is rather easy in our model.

The intention of Chapter 4 was to demonstrate the general benefits of CII processes in the statistical modeling of electricity prices. The exact design of the components of the final price model, however, could be subject to a further statistical analysis. For example, it would be interesting to consider a CII process as the driving component of a more general CARMA process. Regardless of the exact design of the model, the predicted derivative prices should be treated with care. The reason for this is that market participants have access to a variety of different sources for forecasting residual demand. In contrast we only had access to data supported by one provider.

In Chapter 2 and Chapter 4 we have presented different types of electricity price models. In contrast to classical financial mathematics, in which a small number of models dominate, there is no standard methodology for describing electricity prices. This makes model comparisons even more important. Hence, we also compared our models with other approaches to illustrate the benefits. However, this raises the question of the basis on which such an comparison should take place. If accurate predictions are desired, it is logical to use the forecast error as criterion (cf. Weron (2014)). If the focus is on the valuation of derivatives instead, this alone is not always sufficient. Here it can be important to map the characteristics of the data precisely. Often a comparison of moments is used to check this. A finite number of moments, however, provides only limited information. Since the

data is completely characterized by its distribution, the use of information criteria that take into account the density is more reasonable. This includes the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), for example. A comparison of early electricity price models based on such criteria was performed by Bierbrauer et al. (2007). For recent models, such a comparison is more difficult because analytic expressions of the densities are often not available. This is especially true for models that are composed of different processes as it is the case if one accounts for renewable power generation. Here, a numerical approximation of the likelihood as presented in Chapter 2 could be a solution. It would be interesting to investigate this approximation for different electricity price models and to use it as a basis for a profound model comparison.

Appendix A

Source Code

A.1 Implementations for Chapter 2

This section contains the C++ and R code for computing the likelihood $\widehat{L}$ of CTAR(2) with uniform jumps by particle filtering (cf. Chapter 2). The C++ code is placed in two separate files. In "UnweightedBinaryTree.cpp" the unweighted binary tree of Lee (2008) is implemented. The functions in file "CTAR2.cpp" allow to perform parallelized Euler approximations. C++ is connected to R through the package "Rcpp". The individual files are parsed in R by the command **sourceCpp**. The provided C++ functions allow a fast execution of the log likelihood **approxLL**. We also give an example of how to call this function. To get parameter estimates, **approxLL** can be maximized by a numerical optimizer. We used the simultaneous perturbation stochastic approximation (SPSA) algorithm of Spall (2003), which is easy to implement given the objective function.

A.1.1 Content of file "UnweightedBinaryTree.cpp"

```
//[[Rcpp::plugins(cpp11)]]
#include <Rcpp.h>
using namespace Rcpp;

//Node class
struct leafClass{
  double weight;
  int level;
  int index;
  leafClass *left;
  leafClass *right;
```

```
};

//Binary tree class
class btree{
public:
  btree();
  int search(NumericVector u, rootClass *root);
  void destroyTreePublic(rootClass *root);
  rootClass *construct(int dimTree, NumericMatrix particles,
    ↪ NumericVector weights);
private:
  void destroyLeaves(leafClass *leaf);
  void destroyTree(rootClass *root);
  void assignToLeaves(int level, rootClass *root);
  int search(NumericVector u, leafClass *leaf, rootClass *root);
  void construct(leafClass *leaf, rootClass *root);
  rootClass *root;
};

//Initialize binary tree class
btree::btree(){
  root=NULL;
};

//Destroy leaves private
void btree::destroyLeaves(leafClass *leaf){
  if(leaf!=NULL){
    destroyLeaves(leaf->left);
    destroyLeaves(leaf->right);
    delete(leaf);
  }
};

//Destroy tree private
void btree::destroyTree(rootClass *root){
  if(root!=NULL){
    destroyLeaves(root->rootLeaf);
    delete(root);
```

```
  }
};

//Destroy tree public
void btree::destroyTreePublic(rootClass *root){
  destroyTree(root);
};

//Assign data to leaves by median splitting point
void btree::assignToLeaves(int level, rootClass *root){
  //Calculate median
  NumericVector splittingData(pow(2,level));
  int iSplit = 0;
  for(int i=root->splitLevelIndex(root->dimTree-level)*pow(2,
     ↪ level); i<(root->splitLevelIndex(root->dimTree-level)+1)
     ↪ *pow(2,level); i++){
    splittingData(iSplit) = root->particles(root->r,root->
       ↪ particlesIndices(i));
    iSplit += 1;
  }
  double m = median(splittingData);
  //Sort data to leaves
  NumericVector splittingIndices(pow(2,level));
  int iLeft = 0;
  int iRight = pow(2,level-1);
  for(int i=root->splitLevelIndex(root->dimTree-level)*pow(2,
     ↪ level); i<(root->splitLevelIndex(root->dimTree-level)+1)
     ↪ *pow(2,level); i++){
    if(root->particles(root->r,root->particlesIndices(i))<=m){
      splittingIndices(iLeft ) = root->particlesIndices(i);
      iLeft += 1;
    }else{
      splittingIndices(iRight) = root->particlesIndices(i);
      iRight += 1;
    }
  }
  //Assign splitted particles
```

```
  int iAssign = 0;
  for(int i=root->splitLevelIndex(root->dimTree-level)*pow(2,
     ↪ level); i<(root->splitLevelIndex(root->dimTree-level)+1)
     ↪ *pow(2,level); i++){
    root->particlesIndices(i) = splittingIndices(iAssign);
    iAssign += 1;
  }
  //Notice new split index
  root->splitLevelIndex(root->dimTree-level) += 1;
};

//Construct tree private
void btree::construct(leafClass *leaf, rootClass *root){
  //Set splitting dimension
  root->r = (root->dimTree-leaf->level)%root->particles.nrow();
  //Split particles by median
  assignToLeaves(leaf->level, root);
  //Create left leaf
  leaf->left = new leafClass();
  leaf->left->level = leaf->level-1;
  leaf->left->weight = 0;
  for(int i=root->levelIndex(root->dimTree-leaf->left->level)*
     ↪ pow(2,leaf->left->level); i<(root->levelIndex(root->
     ↪ dimTree-leaf->left->level)+1)*pow(2,leaf->left->level);
     ↪ i++){
    leaf->left->weight += root->weights(root->particlesIndices(i
       ↪ ));
  }
  leaf->left->left=NULL;
  leaf->left->right=NULL;
  if(leaf->left->level==0){
    leaf->left->index = root->particlesIndices(root->levelIndex(
       ↪ root->dimTree-leaf->left->level));
  }
  root->levelIndex(root->dimTree-leaf->left->level) += 1;
  //Create right leaf
  leaf->right=new leafClass ();
```

```
  leaf->right->level= leaf->level-1;
  leaf->right->weight = 0;
  for(int i=root->levelIndex(root->dimTree-leaf->right->level)*
     ↪ pow(2,leaf->right->level); i<(root->levelIndex(root->
     ↪ dimTree-leaf->right->level)+1)*pow(2,leaf->right->level)
     ↪ ; i++){
    leaf->right->weight += root->weights(root->particlesIndices(
       ↪ i));
  }
  leaf->right->left=NULL;
  leaf->right->right=NULL;
  if(leaf->right->level==0){
    leaf->right->index = root->particlesIndices(root->levelIndex
       ↪ (root->dimTree-leaf->left->level));
  }
  root->levelIndex(root->dimTree-leaf->right->level) += 1;
  //Normalize weights
  leaf->left->weight = leaf->left->weight/(leaf->left->weight+
     ↪ leaf->right->weight);
  leaf->right->weight = leaf->right->weight/(leaf->left->weight+
     ↪ leaf->right->weight);
  //Construct next level of leaves
  if(leaf->left->level >0){
    construct(leaf->left, root);
    construct(leaf->right, root);
  }
};

//Construct tree public
rootClass *btree::construct(int dimTree, NumericMatrix particles
   ↪ , NumericVector weights){
  //Initialize root
  rootClass * root = new rootClass();
  root->particles = NumericMatrix(particles.nrow(),particles.
     ↪ ncol());
  root->particles = particles;
  root->weights = NumericVector(dimTree);
```

```
  root->weights = weights;
  root->particlesIndices = seq(0,particles.ncol()-1);
  root->levelIndex = NumericVector(dimTree+1);
  root->splitLevelIndex = NumericVector(dimTree+1);
  root->dimTree = dimTree;
  root->rootLeaf=NULL;
  //Initialize root as leaf
  leafClass * rootLeaf = new leafClass();
  rootLeaf->level = dimTree;
  rootLeaf->weight = sum(root->weights);
  rootLeaf->left=NULL;
  rootLeaf->right=NULL;
  //Assign root leaf to root class
  root->rootLeaf = rootLeaf;
  //Construct leaves
  construct(rootLeaf, root);
  //Return root
  return(root);
};

//Search for node private
int btree::search(NumericVector u, leafClass *leaf, rootClass *
   ↪ root){
  //Return leaf index
  if(leaf->level==0){
    return(leaf->index);
  }
  //Set search dimension
  int j = (root->dimTree-leaf->level)%root->particles.nrow();
  //Search for leaf
  if(u(j)<leaf->left->weight){
    u(j) = u(j)/leaf->left->weight;
    return(search(u, leaf->left, root));
  }else{
    u(j) = (u(j)-leaf->left->weight)/(1-leaf->left->weight);
    return(search(u, leaf->right, root));
  }
```

```
};

//Search for node public
int btree::search(NumericVector u, rootClass *root){
  return(search(u, root->rootLeaf, root));
};

//R interface for binary trees
// [[Rcpp::export]]
NumericVector searchConstructTree(int dimTree, NumericMatrix
   ↪ particles, NumericVector weights, NumericMatrix u){
  //Construct binary tree
  class btree Tree;
  rootClass *r = Tree.construct(dimTree, particles, weights);
  //Selecting particles
  NumericMatrix uSearch = NumericMatrix(u.nrow(), u.ncol());
  uSearch = u;
  NumericVector indicesSelect(u.nrow());
  for(int i=0; i<u.nrow(); i++){
    indicesSelect(i) = Tree.search(uSearch(i,_), r);
  }
  //Free memory
  Tree.destroyTreePublic(r);
  return(indicesSelect);
};
```

A.1.2 Content of file "CTAR2.cpp"

```
//[[Rcpp::depends(RcppParallel)]]
#include <RcppParallel.h>
#include <Rcpp.h>
#include <math.h>
#include <numeric>
using namespace Rcpp;

//Data structure for parallel Euler approximation of the state
   ↪ vector of CTAR(2) with uniform jumps
struct StatesParallel : public RcppParallel::Worker{
```

```
//Input parameters
const int index;//Simulation index
const double delta;
const double nEuler;
const int nSim;
const NumericMatrix & par;
const NumericMatrix & diff;
const NumericMatrix & jumpHeights;
const NumericMatrix & jumpTimes;
//Output simulations
RcppParallel::RMatrix<double> output;
//Initialize with source and destination
StatesParallel(const int index, const double delta, const int
   ↪ nSim, const double nEuler, const NumericMatrix &par,
   ↪ const NumericMatrix &diff, const NumericMatrix &
   ↪ jumpHeights, const NumericMatrix &jumpTimes,
   ↪ NumericMatrix output)
  : index(index), delta(delta),nSim(nSim), nEuler(nEuler), par
    ↪ (par), diff(diff), jumpHeights(jumpHeights), jumpTimes
    ↪ (jumpTimes), output(output) {}
//Call function
void operator()(std::size_t begin, std::size_t end) {
  for (int i = begin; i < end; i++) {
    //Initialize Euler
    double auxVar = output.column(i).operator[](0);//Aux.
       ↪ variable
    double a1;//AR parameter vectors
    double a2;
    double sigma = par(4,i);
    double minJump = par(6,i);
    double magnJump = par(7,i);
    double lambda = par(5,i);
    double th = par(8,i);
    double maxJump = minJump+magnJump;
    double pJump = lambda*delta*nEuler;
    for(int j = 0;j<nEuler;j++){
```

```
        if(output(0,i)<th){
          a1 = par(1,i);
          a2 = par(0,i);
        }else{
          a1 = par(3,i);
          a2 = par(2,i);
        }
        //Simulation
        output.column(i).operator[](0) = output.column(i).
            ↪ operator[](0)+delta*output.column(i).operator[](1)
            ↪ ;
        if(jumpTimes((index-1)*nSim+i,j)<=pJump){
          if(jumpHeights((index-1)*nSim+i, j)>=0){
            output.column(i).operator[](1) = output.column(i).
                ↪ operator[](1)+delta*(-a2*auxVar-a1*output.
                ↪ column(i).operator[](1))+sigma*sqrt(delta)*
                ↪ diff((index-1)*nSim+i,j)+(jumpHeights((index
                ↪ -1)*nSim+i,j)*(maxJump-minJump)+minJump);
          }else{
            output.column(i).operator[](1) = output.column(i).
                ↪ operator[](1)+delta*(-a2*auxVar-a1*output.
                ↪ column(i).operator[](1))+sigma*sqrt(delta)*
                ↪ diff((index-1)*nSim+i,j)+(jumpHeights((index
                ↪ -1)*nSim+i,j)*(maxJump-minJump)-minJump);
          }
        }else{
          output.column(i).operator[](1) = output.column(i).
              ↪ operator[](1)+delta*(-a2*auxVar-a1*output.column
              ↪ (i).operator[](1))+sigma*sqrt(delta)*diff((
              ↪ index-1)*nSim+i,j);
        }
        auxVar = output.column(i).operator[](0);
      }
    }
  }
};
```

```
//R interface for parallel Euler approximation of the state
   ↪ vector of CTAR(2) with uniform jumps
//[[Rcpp::export]]
NumericMatrix statesEulerParallel(int index, int nSim, double
   ↪ nEuler, double delta, NumericMatrix &par, NumericMatrix &
   ↪ initStates, NumericMatrix &diff, NumericMatrix &
   ↪ jumpHeights, NumericMatrix &jumpTimes){;
  //Initialize output matrix
  NumericMatrix output(initStates.nrow(),initStates.ncol());
  for(int l=0; l<initStates.nrow(); l++){
    for(int j=0; j<nSim; j++){
      output(l,j) = initStates(l,j);
    }
  }
  //Initialize structure for parallel Euler
  struct StatesParallel statesParallel(index, delta, nSim,
     ↪ nEuler, par, diff, jumpHeights, jumpTimes, output);
  //Run parallel Euler
  parallelFor(0, nSim, statesParallel);
  //Return
  return(output);
};
```

A.1.3 R code for approximate likelihood

```
library(Rcpp)
library(RcppParallel)
sourceCpp("...//UnweightedBinaryTree.cpp")
sourceCpp("...//CTAR2.cpp")

### Approximate likelihood function of CTAR(2)
### with uniform jumps (cf. Equation (2.30))

approxLL <- function(par, Y, nSim, nEuler, delta, diffusion,
   ↪ jumpHeights, jumpTimes, crnTree){
  ###Initialize particle filter
  statesFilter <- matrix(0, 2, nSim)
  statesFilter[1,] <- Y[1]
  bwResampling <- rep(0, 2)
```

```
particles <- matrix(0, 2, nSim)
ll <- 0
statesFilter <- matrix(0, nrow=p, ncol=nSim)
###Initialize parallel Euler
parMatrix <- matrix(par, length(par), nSim)
initStates <- matrix(0, 2, nSim)
###Particle Filter
for(i in 1:(length(Y)-1)){
  ###Parallel Euler
  initStates <- statesFilter
  simStates <- statesEulerParallel(i, nSim, nEuler, delta,
    ↪ parMatrix, initStates, diffusion, jumpHeights,
    ↪ jumpTimes)
  ###Calculate likelihood
  bw <- 1.06*sd(simStates[1,])/(nSim^(1/5))
  dens <- 1/nSim*sum(dnorm(Y[i+1], simStates[1,], bw))
  ll <- ll+log(dens)
  ###Resampling particles
  particles[1,] <- Y[i+1]
  particles[2,] <- simStates[2,]
  ###Resampling weights
  for(l in 1:2){
    bwResampling [l] <- 1.06*sd(simStates[l,])/(nSim^(1/5))
  }
  resamplingWeights <- dnorm(simStates[1,], Y[i+1],
    ↪ bwResampling[1])
  ###Binary tree resampling
  normWeights <- resamplingWeights/sum(resamplingWeights)
  dimTree <- log(nSim,2)
  uVec <- crnTree[seq((i-1)*nSim+1,i*nSim), ]
  resamplingIndices <- searchConstructTree(dimTree, particles,
    ↪ normWeights,uVec)+1
  statesFilter[1,] <- Y[i+1]
  statesFilter[2,] <- particles[2, resamplingIndices]
}
return(ll)
```

```
}
###
# Example of a function call of approxLL
###


###Observations
Y <- rnorm(100)
nData <- length(Y)

###Simulation parameters
nSim <- 2^11 #Number of simulated particles
nEuler <- 100 #Number of Euler steps
delta <- 1/100 #Euler stepsize

###CTAR(2) with uniform jumps parameters
sigma <- 0.6 #Diffusion standard deviation
lambda <- 0.05 #Jump intensity
minJump <- 0.25 #Minimal jump size
magJump <- 0.77 #Jump magnitude
th <- 2.2 #Threshold
a11 <- 1 #AR parameters
a12 <- 0.2
a21 <- 2
a22 <- 0.6
par <- c(a12, a11, a22, a21, sigma, lambda, minJump, magJump, th
  ↪ )

###Simulate stochastic components
diffusion <- matrix(0, nData*nSim, nEuler) #Increments Brownian
  ↪ motion
jumpHeights <- matrix(0, nData*nSim, nEuler) #Jump sizes
jumpTimes <- matrix(0, nData*nSim, nEuler) #Jump times
crnTree <- matrix(0, 2*nData*nSim, 2) #Random number binary tree
  ↪ indices
for(col in 1:nEuler){
  diffusion[,col] <- rnorm(nData*nSim)
```

```
  jumpsSign <- c(1,-1)
  jumpHeights[,col] <- runif(nData*nSim)*sample(jumpsSign, nData
    ↪ *nSim, TRUE)
}
jumpTimes <- matrix(runif(nData*nSim*nEuler),nData*nSim,nEuler)
for(col in 1:2){
  crnTree[,col] <- runif(2*nData*nSim)
}

###Call likelihood function
approxLL(par, Y, nSim, nEuler, delta, diffusion, jumpHeights,
  ↪ jumpTimes, crnTree)
```

A.2 Implementations for Chapter 3

Here we give R functions to calculate the solar angle θ^{loc} of an object depending on date, position and tilt. This is used in Chapter 3 to determine the attenuation coefficients μ_i, $i = 1, \ldots, N$ (cf. Equation (3.7)).

```
### Solar angle in radians for a given date and position
solarAngle <- function(month, day, hour, min, lat, long){
        K <- pi/180
        dayNumber <- (month-1)*30.3+day
        declin <- -23.45*cos(K*360*(dayNumber+10)/365)
        timeEquation <- 60*(-0.171*sin(0.0337*dayNumber + 0.465)
          ↪ - 0.1299*sin(0.01787*dayNumber - 0.168))
        hourAngle <- 15*(hour + min/60 - (15.0-long)/15.0 - 12 +
          ↪ timeEquation/60)
        alt <- sin(K*lat)*sin(K*declin)+cos(K*lat)*cos(K*declin)
          ↪ *cos(K*hourAngle)
        angle <- asin(alt)/K
        return(angle)
}

### Object's solar angle in radians for a given tilt, date and position
objAngle <- function(tilt, month, day, hour, min, lat, long){
```

```
    zenithAngle <- solarAngle(month, day, hour, min, lat,
      ↪ long)-90
    zenithAngle[zenithAngle<=0] <- NA
    zenithAngle[zenithAngle>0] <- zenithAngle[zenithAngle
      ↪ >0]+tilt
    return(zenithAngle)
}
```

A.3 Implementations for Chapter 4

In the following we provide an R implementation of the nonparametric moment estimator $\widehat{m}_{k,n}$ and the test statistic S_n introduced in Chapter 4. In addition, the objective functions for specifying the bandwidth parameter ϵ_n are given. To find an optimal bandwidth, these can be put into a suitable numerical optimizer. We used the Nelder–Mead method implemented in the R function **optim**.

Kernel function

```
kernel <- function(x, mean, bw, type){
  if(type=="epanechnikov"){
    kernel <- 3/4*(1-((x-mean)/bw)^2)*(abs((x-mean)/bw)<=1)
  }
  if(type=="normal"){
    kernel <- dnorm(x, mean=mean, sd=bw)
  }
  if(type=="unif"){
    kernel <- (abs((x-mean)/bw)<=1)
  }
  return(kernel)
}
```

Nonparametric moment estimator (cf. Equation (4.4))

```
npMoment <- function(x, k, I, DeltaX, eps, ktype="normal"){
  invalid <- is.na(DeltaX)
  est <- sum(kernel(I[!invalid], x, eps, ktype)*DeltaX[!invalid
    ↪ ]^k)/sum(kernel(I[!invalid], x, eps, ktype))
  return(est)
```

```
}
### Objective function for LSCV bandwidth selection (cf. Equation (4.7))
lscvObj <- function(eps, k, I, DeltaX, ktype="normal"){
  n <- length(I)
  invalid <- is.na(DeltaX)
  seqValid <- seq(1,n)[!invalid]
  momEst <- rep(0,length(seqValid))
  j <- 1
  for(i in seqValid){
    momEst[j] <- npMoment(I[i], k, I[-i], DeltaX[-i], eps, ktype
      ↪ )
    j <- j+1
  }
  ls <- sum((DeltaX[!invalid]^k-momEst)^2)
  return(ls)
}
### Statistic for testing the null hypotheses of constant moments (cf. Equation (4.8))
testStatistic <- function(evalPoints, k, I, DeltaX, eps, ktype="
  ↪ unif"){
  invalid <- is.na(DeltaX)
  mkConst <- mean(DeltaX[!invalid]^k)
  m2kConst <- mean(DeltaX[!invalid]^(2*k))
  mk <- apply(t(evalPoints), 2, npMoment, k, I, DeltaX, eps,
    ↪ ktype)
  Nx <- colSums(apply(t(evalPoints), 2, kernel, I[!invalid], eps
    ↪ , ktype))
  Sn <- 1/(m2kConst-mkConst^2)*sum(N_x*(mk-mkConst)^2)
  return(Sn)
}
### Objective function for selecting test bandwidth parameter
tuningTestObj <- function(eps, evalPoints, k, I, DeltaLevyMatrix
  ↪ , ktype="unif"){
  nTests <- dim(DeltaLevyMatrix)[1]
  Sn <- rep(0,nTest)
  for(iTest in 1:nTest){
```

```
    S_n[iTest] <- testStatistic(evalPoints, k, I,
      ↪ DeltaLevyMatrix[iTest,], eps, ktype)
  }
  alpha <- seq(0,1,0.01)
  rejectionRate <- rep(0,length(alpha))
  for(iAlpha in 1:length(alpha)){
        cValue <- qchisq(1-alpha[iAlpha],df=length(evalPoints))
        rejectionRate[iAlpha] <- sum(Sn>=cValue)/length(Sn)
  }
  error <- sum((rejectionRate-alpha)^2)
  return(error)
}
```

Abbreviations

a.s.	almost surely
ACF	autocorrelation function
ACVF	autocovariance function
AGEB	Arbeitsgemeinschaft Energiebilanzen e. V.
AIC	Akaike information criterion
BIC	Bayesian information criterion
BNetzA	Bundesnetzagentur
CAR	continuous-time autoregressive
CARMA	continuous-time autoregressive moving average
CII	conditionally independent increment
CTAR	continuous-time threshold autoregressions
EEG	Gesetz für den Ausbau erneuerbarer Energien (Renewable Energy Sources Act)
EEX	European Energy Exchange
ENTSO-E	European Network of Transmission System Operators
EnWG	Gesetz über die Elektrizität- und Gasversorgung (Law on Energy Management)
EPEX SPOT	European Power Exchange
Eurostat	European Statistical Office
GH	generalized hyperbolic
GW	gigawatt

i.i.d.	independent and identically distributed
KS	Kolmogorow-Smirnow
LSS	regime-switching Lévy semistationary
MCP	market clearing price
MW	megawatt
MWh	megawatt hour
NIG	normal-inverse Gaussian
OU	Ornstein-Uhlenbeck
PACF	partial autocorrelation function
Phelix	physical electricity index
PV	photovoltaic
SDE	stochastic differential equation
SPSA	simultaneous perturbation stochastic approximation
TWh	terawatt hour

List of Symbols

$\approx$	approximately
$\xrightarrow{d}, \xrightarrow{\mathcal{L}}$	convergence in distribution
$\xrightarrow{L^n(\Omega)}$	convergence in the n-th mean
$\xrightarrow{p}, \xrightarrow{\mathbb{P}}$	convergence in probability
$\xrightarrow{a.s.}$	almost sure convergence
$\overset{i.i.d}{\sim}$	independent and identically distributed as
$\sim$	is distributed as
$\overset{d}{=}$	equally distributed
$\perp\!\!\!\perp$	independent
$\propto$	directly proportional

$ber(p)$	Bernoulli distribution with parameter p
$bin(n, p)$	binomial distribution with parameters n and p
$\phi(\cdot)$	density of a standard normal distribution
$\Phi(\cdot)$	standard normal cumulative distribution function
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$pois(\lambda)$	Poisson distribution with parameter n λ
$\mathcal{WN}$	independent with mean zero and finite variance (white noise)

argmax	argument of the maximum
$\circ$	function composition
$\cos(\cdot)$	cosines function
$(\cdot)'$	derivative

$\|\cdot\|$	Euclidean norm
$\exp(\cdot)$	exponential function
$\lfloor\cdot\rfloor$	floor function
$\mathbf{1}_X, \mathbf{1}(X)$	indicator function for a set X
inf	infimum
lim sup	limes superior
$\text{logit}(\cdot)$	logit function
$\log(\cdot)$	logarithmic function
max	maximum
$X \wedge Y$	minimum of X and Y
$\text{Re}(\cdot)$	real part
$\sec(\cdot)$	secant function
$\sin(\cdot)$	sinus function
sup	supremum
$(\cdot)^T$	transposition
$\mathbb{E}(\cdot\|X)$,$\mathbb{E}(\cdot\|\mathcal{F})$	conditional expectation with respect to a random variable X resp. a σ-algebra $\mathcal{F}$
$\gamma_X(\cdot)$	autocovariance function of a random variable X
$Cov(\cdot)$	covariance matrix
$Cov(\cdot,\cdot)$	covariance function
$\mathbb{E}_{\mathbb{P}}(\cdot)$	expectation with respect to $\mathbb{P}$
$\mathbb{E}(\cdot)$	expectation
$[\cdot,\cdot]$	covariation
$[\cdot]$	quadratic variation
$Var(\cdot)$	variance
$\overline{E}, E^c$	complement of a set E
X^c	continuous part of X
€	Euro
I_n	$n \times n$ identity matrix
I^{-1}	inverse of a matrix I
Σ	covariance matrix
$\mathbf{1}_n$	n-th canonical unit vector

Ω	sample space
$\mathcal{F}$	σ-algebra
$\mathbb{P}(\cdot\|\mathcal{F})$,$\mathbb{P}_{\|A}(\cdot)$,$\mathbb{P}(\cdot\|A)$	conditional probability measure on the σ-algebra $\mathcal{F}$ resp. the event A
$\mathcal{L}_{\|A}(X)$,$\mathcal{L}(X\|A)$	law of X under the event A
$\mathcal{L}(X)$	law of X
$\mathbb{F}$	filtration
$\mathcal{F}^X$	sigma-algebra generated by X
ω	probability event
$\mathbb{P}(\cdot)$	probability measure
$B_r(x)$	closed 1-ball of radius r and center x
$\mathcal{B}(E)$	Borel algebra on E
$\mathbb{C}$	set of all complex numbers
$C_b^n(E)$	set of continuous functions on E with bounded continuous first n derivatives
$C(E), C(E,M)$	set of real- resp. M-valued continuous functions on E
$\mathcal{O}(f)$	set of functions with the same asymptotic behavior as f
$\mathbb{N}$	set of all natural numbers
$C^n(E)$	set of continuous functions on E with continuous first n derivatives
$\mathbb{R}^+$	set of all positive real numbers
$\mathbb{R}$	set of all real numbers
$D(E,M)$	set of all càdlàg functions from E to M

Bibliography

Arbeitsgemeinschaft Energiebilanzen e.V. (2017). Energieverbrauch in Deutschland im Jahr 2017.

Arbeitsgemeinschaft Energiebilanzen e.V. (2018). Auswertungstabellen zur Energiebilanz Deutschland.

Bacher, P., H. Madsen, and H. A. Nielsen (2009). Online short-term solar power forecasting. *Solar Energy 83*, 1772–1783.

Barlow, M. (2002). A diffusion model for electricity prices. *Mathematical Finance 12*, 287 – 298.

Barndorff-Nielsen, O. E. (1997). Normal inverse gaussian distributions and stochastic volatility modelling. *Scand. J. Statist. 24*, 1–13.

Barndorff-Nielsen, O. E. and N. Shepard (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal of Royal Statistical Society. Series B 63*(2), 167–241.

Benth, F. E., J. S. Benth, and S. Koekebakker (2008). *Stochastic Modelling of Electricity and Related Markets*. Singapore: World Scientific.

Benth, F. E. and N. A. Ibrahim (2017). Stochastic modeling of photovoltaic power generation and electricity prices. *Journal of Energy Markets 10*, 1–33.

Benth, F. E., C. Klüppelberg, G. Müller, and L. Vos (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energy Economics 44*, 392–406.

Benth, F. E. and A. Pircalabu (2018). A non-gaussian ornstein-uhlenbeck model for pricing wind power futures. *Applied Mathematical Finance 25*(1), 36–65.

Benth, F. E. and J. Saltyte Benth (2011). Weather derivatives and stochastic modelling of temperature. *International Journal of Stochastic Analysis*. Vol. 2011, Article ID 576791, 21 pages.

Benth, F. E., J. Saltyte Benth, and S. Koekebakker (2007). Putting a price on temperature. *Scand. J. Stat. 34*, 746–767.

Bierbrauer, M., C. Menn, S. T. Rachev, and S. Trück (2007). Spot and derivative pricing in the eex power market. *Journal of Banking and Finance 31*, 3462 – 3485.

Billingsley, P. (1999). *Convergence of Probability Measures* (2 ed.). Wiley, New York.

Borovkova, S. and F. J. Permana (2006). Modelling electricity prices by the potential jump-diffusion. In A. Shiryaev, M. Grossinho, P. Oliveira, and M. Esquivel (Eds.), *Stochastic Finance*. Springer, Bosten, MA.

Brockwell, P. J. (1994). On continuous-time threshold ARMA processes. *Journal of Statistical Planning and Inference 39*, 291–303.

Brockwell, P. J. (2001a). Continuous-time ARMA processes. In D. N. Shanbhag and R. C. R. (Eds.), *Handbook of statistics 19*, pp. 249–276. North-Holland.

Brockwell, P. J. (2001b). Levy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics 53*, 113–124.

Brockwell, P. J. (2014). Recent results in the theory and application of CARMA processes. *Annals of the Institute of Statistical Mathematics 66*(4), 647–685.

Brockwell, P. J., R. A. Davis, and Y. Yu (2007). Continuous-time gaussian autoregression. *Statistica Sinica 17*, 63–80.

Brockwell, P. J. and R. J. Hyndman (1992). On continuous-time threshold autoregression. *International Journal of Forecasting 8*, 157–173.

Brockwell, P. J. and A. Lindner (2015). Prediction of stationary lévy-driven CARMA processes. *Journal of Econometrics 189*(2), 263–271.

Brockwell, P. J. and R. J. Williams (1997). On the existence and application of continuous-time threshold autoregressions of order two. *Advances in Applied Probability 29*, 205–227.

Bundesnetzagentur (2018). Monitoringbericht 2018.

Bundesnetzagentur (2019). SMARD Strommarktdaten. (as of 31/01/2019). `http://www.smard.de`.

Burger, M., B. Klar, A. Müller, and G. Schindlmayer (2004). A spot market model for pricing derivatives in electricity markets. *Quantitative Finance 4*(1), 109 – 122.

Campillo, F. and V. Rossi (2009). Convolution particle filtering for parameter estimation in general state-space models. *IEEE Transactions on Aerospace and Electronic Systems 45*(3), 1063–1072.

Chopin, N. (2004). Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *Ann. Stat. 32*(6), 2385–2411.

Cont, R. and P. Tankov (2004). *Financial Modelling With Jump Processes*. Chapman & Hall.

Deschatre, T. (2017). *Dependence modeling between continuous time stochastic processes: an application to electricity markets modeling and risk management*. Ph. D. thesis, PSL Research University.

Deschatre, T. and A. Veraart (2017). A joint model for electricity spot prices and wind penetration with dependence in the extremes. Available at SSRN:

https://ssrn.com/abstract=3029318.

Doob, J. L. (1942). The brownian movement and stochastic equations. *Annals of Mathematics 43*(2), 351–369.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2 ed.). Cambridge University Press.

Ehrlich, E., A. Jasra, and N. Kantas (2015). Gradient free parameter estimation for hidden markov models with intractable likelihoods. *Methodology and Computing in Applied Probability 17*(3), 315–349.

Eriksson, A., E. Ghysels, and W. Fangfang (2009). The normal inverse gaussian distributions and the pricing of derivatives. *The Journal of Derivatives 16*(3), 23–37.

Ethier, S. N. and T. G. Kurtz (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

European Network of Transmission System Operators (2019). Transparency plattform. (as of 31/01/2019).
`https://transparency.entsoe.eu/dashboard/show`.

European Statistical Office (2019, NOTE =). Renewable energy statistics.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient form discrete observations. *Journal of Applied Probability 30*(4), 790–804.

Föllmer, H. and P. Protter (2011). Local martingales and filtration shrinkage. *ESAIM Probability and Statistics 15*, 25–38.

Garcia, I., C. Klüppelberg, and G. Müller (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Statistical Modelling 11*(5), 447–470.

Gourieroux, C. and A. Monfort (1996). *Simulation Based Econometric Methods*. Oxford University Press, Oxford.

Graeber, D. R. (2014). *Handel mit Strom aus erneuerbaren Energien*. Springer Gabler.

Grigelionis, B. (1975). The characterization of stochastic processes with conditionally independent increments. *Litovsk. Math. Sb. 15*, 53–60.

Honoré, P. (1998). Pitfalls in estimationg jump-diffusion models. Technical report, University of Aarhus Graduate School of Business.

Hyndman, R. H. (1992, 12). *Continuous time threshold autoregressive models*. Ph. D. thesis, University of Melbourne.

Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *Scandinavian Journal of Statistics 27*(1), 83–96.

Jacod, J. (2004). The euler scheme for lévy driven stochastic differential equations: limit theorems. *The Annals of Probability 32*(3), 1830–1872.

Jacod, J., T. Kurtz, S. Méléard, and P. Protter (2005). The approximate euler method

for levy driven stochastic differential equations. *Ann. Inst. H. Poincare, Probab. Statist. 41*(3), 523–558.

Jacod, J. and P. Protter (2012). *Discretization of Processes*. Springer-Verlag Berlin Heidelberg.

Jacod, J. and A. N. Shiryaev (2002). *Limit theorems for Stochastic Processes* (2 ed.). Springer-Verlag, Berlin Heidelberg.

Karatzas, I. and S. E. Shreve (1998). *Brownian Motion and Stochastic Calculus* (2 ed.). Springer Science+Business Media, New York.

Klebaner, F. and R. Lipster (2011). When a stochastic exponential is a true martingale. extension of the beneš method. *Theory of Probability and Its Applications 58*(1), 38–62.

Klenke, A. (2008). *Wahrscheinlichkeitstheorie* (2 ed.). Springer-Verlag Berlin Heidelberg.

Kurtz, T. G. and P. Protter (1991). Weak limit theorems for stochastic integrals and stochastic differntial equations. *The Annals of Probability 19*(3), 1035–1070.

Lee, A. (2008). Towards smoother multivariate particle filters. Master's thesis, University of British Columbia, Vancouver, BC.

Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics:Theory and Practice*. Princeton University Press.

Lingohr, D. and G. Müller (2019). Stochastic modelling of intraday photovoltaic power generation. *Energy Economics*. https://doi.org/10.1016/j.eneco.2019.03.007.

Malik, S. and M. Pitt (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics 165*(2), 190–209.

Maller, R. A., G. Müller, and A. Szimayer (2009). Ornstein-uhlenbeck processes and extensions. In T. Mikosch, J.-P. Kreiß, R. A. Davis, and T. G. Andersen (Eds.), *Handbook of Financial Time Series*, pp. 421–437. Springer-Verlag Berlin Heidelberg.

Mueller, R. W., K. F. Dagestad, P. Ineichen, M. Schroedter-Homscheidt, S. Cros, and D. Dumortier (2004). Rethinking satellite-based solar irradiance modelling: The solis clear-sky module. *Remote Sensing of Environment 91*, 160–174.

Pedro, H. T. C. and C. F. M. Coimbra (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy 86*, 2017–2028.

Pitt, M., R. Silva, P. Giordani, and R. Kohn (2012). On some properties of markov chain monte carlo simulation methods based on the particle filter. *Journal of Econometrics 171*(2), 134–151.

Pitt, M. K. (2002). Smooth particle filters for likelihood evaluation and maximisation. Warwick Economic Research Papers, University of Warwick.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations* (2 ed.). Springer-Verlag Berlin Heidelberg.

Ren21 (2017). Renewables 2017 global status report. *Paris, REN21 Secretariat*.

Rintamäki, T., A. Siddiqui, and A. Salo (2017). Does renewable energy generation decrease the volatility of electricity prices? an analysis of denmark and germany. *Energy Economics 62*, 270–282.

Rossi, V. and J. Vila (2006). Nonlinear filtering in discrete time: A particle filtering approach. *Publ. Inst. Stat. Univ. Paris* (3), 71–102.

Samuelson, P. (1965). Proof that properly anticipating prices fluctuate randomly. *Indust. Manag. Rev. 6*(2), 41–49.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.

Schlemm, E. and R. Stelzer (2012). Quasi maximum likelihood estimation for strongly mixing state space models and multivariate lévy-driven CARMA processes. *Electron. J. Stat. 6*, 2185–2234.

Schmeck, M. D. and S. Borovkova (2017). Electricity price modeling with stochastic time change. *Energy Economics 63*, 51–65.

Spall, J. C. (2003). *Introduction to stochastic search and optimization*. Wiley, New York.

Stramer, O., R. L. Tweedie, and P. J. Brockwell (1996). Existence and stability of continuous time threshold ARMA processes. *Statistica Sinica 6*, 715–732.

Tong, H. (1983). Threshold models in non-linear time series analysis. In *Springer Lecture Notes in Statistics 21*. Springer-Verlag, New York.

Tómasson, H. (2015). Some computational aspects of gaussian CARMA modelling. *Stat. Comput. 25*, 375–387.

Uhlenbeck, G. E. and L. S. Ornstein (1930). On the theory of brownian motion. *Phys. Rev. 36*, 823–841.

Van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*(2), 177–188.

Veraart, A. E. D. (2016). Modelling the impact of wind power production on electricity prices by regime-switching lévy semistationary processes. In F. E. Benth and G. Di Nunno (Eds.), *Stochastics of Environmental and Financial Economics*, pp. 321–340. Springer Proceedings in Mathematics & Statistics.

Veraart, A. E. D. and H. Zdanowicz (2016). Modelling and predicting photovoltaic power generation in the eex market. Available at SSRN: hhps://ssrn.com/abstract=2691906.

Vila, J. (2012). Enhanced consistency of the resampled convolution particle filter. *Statistics and Probability Letters 82*, 786–797.

Wagner, A. (2012). Residual demand modeling and application to electricity pricing. *The Energy Journal 35*(2), 45–73.

Weron, R. (2006). *Modeling and forecasting electricity loads and prices: a statistical approach*. Wiley.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting 30*, 1030–1081.

Yan, L. (2002). The euler scheme with irregular coefficients. *The Annals of Probability 30*(3), 1172–1194.

Yang, H., M. Hildmann, F. Herzog, and G. Andersson (2013). Modeling the merit order curve of the european energy exchange power market in germany. *IEEE Transactions on power systems 28*(3), 3155–3164.

Äit Sahalia, Y. and J. Jacod (2014). *High Frequency Financial Econometrics*. Princeton University Press.