Stochastic modelling and statistical inference for electricity prices, wind energy production and wind speed

a thesis presented for the degree of

Doctor of Philosophy of Imperial College London

and the

Diploma of Imperial College

by

Paulina A. Rowińska

Department of Mathematics

Imperial College

180 Queen's Gate, London SW7 2BZ

March 2020

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed: _____

Copyright

Thesis advisors: Almut Veraart, Olivier Féron Paulina A. Rowińska

Stochastic modelling and statistical inference for electricity prices, wind energy production and wind speed

Abstract

Although wind energy helps us slow down the increase of global temperatures, its weather-dependence and unpredictability make it risky to invest in. In this thesis we apply statistical and mathematical tools to enable energy providers to accurately plan such investments.

In the first part we want to understand the impact of wind energy on electricity prices. We extend an existing multifactor model of electricity spot prices by including stochastic volatility as well as the information about wind energy production. Empirical studies indicate that these additions improve the model fit. We also model wind-related variables directly, using Brownian semistationary processes with generalised hyperbolic marginals. Finally, we introduce a joint model of prices and wind energy production suitable for quantifying the risk faced by energy distributors.

The second goal is to produce accurate short-term wind speed forecasts based on historical data instead of computationally expensive physical models. We achieve this by splitting the wind speed into two horizontal components and modelling them with Brownian semistationary processes with a novel triple-scale kernel. We develop efficient estimation and forecasting procedures. Empirical studies show that such modelling choices result in good forecasting performance.

To Grandpa Mietek, my greatest supporter. I miss you!

Acknowledgements

Some say that PhD is a painful and lonely journey. While it did feel so at times, I consider myself extremely lucky to have been surrounded by supportive and caring people.

First and foremost, I would like to show my greatest appreciation to my supervisor Almut Veraart. I quickly learned that research barely ever goes according to plan – but thanks to Almut's involvement I was never left alone with my struggles. She cared about much more than just research: she was genuinely interested in my life and me as a person, not just a PhD candidate. I am grateful for her understanding and support in good and bad times: when I needed additional time for science communication activities, but also when everything in my life seemed to fall apart. Almut's kindness, professionalism and reliability made her the best supervisor I could possibly imagine.

I would also like to thank Pierre Gruet, Olivier Ferón and Peter Tankov for fruitful collaboration and invaluable insights into the world of energy industry. I am especially grateful to Pierre for stepping in anytime Almut was not available.

Thanks to my examiners, Emma McCoy and Gernot Müller, my PhD viva was a pleasant and enriching experience. I would like to thank them for valuable comments which greatly improved my thesis.

This research would not have been possible without the financial support from EPSRC under the grant EP/L016613/1 as well as from EDF, which I am very grateful for.

Software tends to give up in most crucial moments. Luckily, no technical issue has been left unsolved thanks to José and Oscar, to whom I am greatly indebted.

Mathematicians have a reputation of being antisocial and boring. My experience is completely different: the Department of Mathematics at Imperial College became my second home, thanks to friends and colleagues I had the

privilege to meet during my PhD.
In particular, turning up at the office every single morning would have been much more challenging without my inspiring office mates. I would like to thank Dorian, Josie, Melissa and Mike for sharing with me the ups and downs of PhD life. I hope they can forgive me for my loud complaints about disappointing simulation results and turning the office into SIAM hub for a year. I am particularly lucky to have been assigned a desk next to Melissa, who quickly became my close friend.
The perspective of sharing the latest gossip over a plate of SCR lunch made even the hardest mornings bearable. I am grateful to members of the Inclusive Food Group and fifth floor stats PhD students for keeping me company.
I will always remember 2017/18 as the year of SIAM. I am grateful to Andreas, Hemant, Johannes, Marco, Tasmin and Yibei for their crazy ideas, enthusiasm and willingness to help even in most unusual tasks.
Finally, I want to thank the people who believed in me more than I believed in myself.
I feel privileged to have shared a big chunk of my PhD journey with Marco, the most encouraging and positive person I have ever met. I am grateful to him for forcing me to take breaks at work, cheering me up with increasingly ridiculous gifts and somehow mastering the task of turning my tears into laughter.
And of course, no words can express my gratitude to my family who have unconditionally supported me from the day I was born. Mum, Dad, Grandma Ala, Grandpa Zygmunt, Grandma Halina and, dearly missed, Grandpa Mietek: I would not have achieved this without you. Thank you for showing me that dreams are worth fighting for!

The answer, my friend, is blowin' in the wind
The answer is blowin' in the wind
Bob Dylan (1962)

List of Figures

List of Tables

Contents

1 Introduction

1.1 Motivation

In December 2015 for the first time governments from all over the world united to tackle climate change. The Paris Agreement helped set concrete and actionable goals to stop the increase in global average temperature before it reaches 2 °C above pre-industrial levels ([United Nations, 2015]). A necessary step to accomplish this ambitious objective is a more effective use of renewable energy sources. In particular, by 2030 the European Union aims to increase the participation of renewable sources in the total energy production to at least 32%, while the energy efficiency by at least 32.5% ([European Commission, 2018a]).

One of the most important renewable energy sources is wind power. Currently it accounts for over 11% of total energy produced in Europe and by 2030 we can expect this number to grow up to 25% ([Wind Europe, 2019]). In the past two decades global wind energy capacity expanded rapidly, from 7.5 GW in 1997 to 564 GW in 2018 ([International Renewable Energy Agency, 2018]). Wind power's dependence on weather conditions makes it notoriously risky to invest in, since weather forecasts, especially with long horizons, still lack the

desirable level of accuracy ([Met Office, 2019]). Therefore energy providers face two important challenges. Firstly, they have to accurately predict the output of wind turbines, which enables them to plan their production and further investments accordingly. Secondly, to assess their potential profit they must be able to understand the impact of wind energy on electricity prices. Our research was motivated by these two questions, which we introduce in more detail in the introductions to Part I and Part II. The practitioners' feedback enabled us to make sure that our results not only advance the mathematical and statistical knowledge but also can be directly applied to real world problems.

1.2 Outline

In Part I we explore the link between electricity prices and wind energy using the Austrian-German market as an example. We start by analysing the historical data in Chapter 3, with particular focus on the relationship between wind-related variables and spot prices of energy. In Chapter 4 we use this knowledge to propose a model of electricity prices with a novel feature: the information about wind energy production. Our empirical results clearly state that such approach improves the model fit and should be considered by practitioners. We explore this idea even further in Chapter 5, where we present a joint model of wind energy production and electricity prices. In building this model we focus on risk quantification, a feature of great importance for energy distributors and traders.

Another challenge faced by the wind energy industry is the shortage of accurate wind speed forecasts. Therefore in Part II we take another approach to help energy providers reach the goal of increased global wind energy production. We aim to accurately predict wind speeds in a given location on short time horizon, which directly translates to short-term wind energy production forecasts. We start by analysing the historical wind speed and direction data from a location near Paris in Chapter 7. Our findings suggest modelling two horizontal components of wind speed separately, which we successfully achieve in Chapter 8. In Chapter 9 we explore the correlation structure of these

components and ways of extending our results to a two-dimensional model. We show that short-term wind speed forecasts based on the two-component model outperform not only simple persistence forecasts but also direct wind speed forecasts (without the horizontal component split).

1.3 Main contributions

As the title of the thesis suggests, our research focused on applying methods of stochastic modelling and statistical inference to increase the knowledge about electricity prices, wind energy production and wind speed forecasting. In this section we list our main contributions to the field described in both parts of the thesis.

In Part I:

- We generalised the arithmetic model for electricity prices proposed by [Benth et al., 2014] so that the short-term factor can accommodate stochastic volatility.
- We extended the model by including the information about wind energy generation, which improved the model fit.
- We applied the Brownian semistationary framework with generalised hyperbolic marginals and gamma kernel introduced by [Barndorff-Nielsen et al., 2013] to model wind-related variables directly.
- We also proposed a discrete joint model of wind energy generation and electricity prices particularly suitable to quantify risks faced by wind energy distributors.

In Part II:

- We modelled wind speed components using Brownian semistationary processes and proposed a method of joining two correlated components into the wind speed.

- We generalised more standard gamma and power-law kernels and proposed a triple-scale kernel able to capture the complex behaviour of wind speed components.
- We studied two methods of parameter estimation: the generalised method of moments and a novel regression-based method. We provided practitioners with recommendations for the estimation procedure.
- We concluded from empirical studies that the split into two horizontal components improves the forecasts in comparison to modelling wind speed directly.

1.4 Papers

The research was done in collaboration with Almut Veraart (Imperial College London), Olivier Féron (EDF), Pierre Gruet (EDF) and Peter Tankov (ENSAE ParisTech). It has resulted in the following papers.

- A paper "A multifactor approach to modelling the impact of wind energy on electricity spot prices" based on Chapter 4, co-authored by Almut Veraart and Pierre Gruet, is currently under revision for resubmission in *Energy Economics*. The preprint is available online: [Rowińska et al., 2018].
- Chapter 5 will result in a paper "A joint model of wind energy production and electricity spot prices" co-authored by Almut Veraart.
- Part II is a foundation of a working paper "Gaussian moving average processes applied to wind speed modelling and forecasting" co-authored by Peter Tankov and Almut Veraart.

Part I

Modelling electricity prices and their relationship with wind energy

2

Introduction

2.1 Motivation

Wind energy constitutes an increasingly higher part of total energy generation all over the world, which is changing the behaviour of energy markets. The combination of its low marginal costs and the inelasticity of power demand (Section 2.2) results in lower energy prices, especially when the wind is high. We call this phenomenon a *merit-order effect* ([European Wind Energy Association, 2010]). It is extremely important to quantify these effects using appropriate mathematical and statistical models. While other renewable sources also impact the supply curve, their impact is more predictable. Therefore in our research we focus on wind energy.

While the literature abounds with models of electricity prices – spot and futures via spot (eg. [Carmona et al., 2013], [Cartea et al., 2009], [Benth et al., 2014]) and futures directly (eg. [Benth and Paraschiv, 2016], [Barndorff-Nielsen et al., 2013], [Borovkova and Geman, 2006]) – most of them ignore the increasing role of renewables, in particular wind power, in energy markets. Few exceptions include [Elberg and Hagspiel, 2015], who included a copula model for the spatial dependence structure of wind power in Germany into

a model for electricity spot prices. [Veraart, 2016] took a different approach and modelled the impact of wind power generation on German spot prices by regime-switching Lévy semistationary processes. [Ketterer, 2014] chose a generalised autoregressive conditional heteroskedasticity (GARCH) model of wind power's impact on the electricity price level and volatility in Germany, taking into account changes in market regulations. Finally, [Deschatre and Veraart, 2017] focused on the impact of wind energy production on the spikes in the spot prices.
In the first part of this thesis we aim to help fill this research gap. We start by describing the main features of electricity prices and introducing the most important energy contracts.

2.2 Features of electricity prices

Because energy markets have been liberalised only over the last decades, modelling electricity prices is a relatively new topic in mathematics and economics. Electricity differs from other commodities ([ter Haar, 2010]), so we cannot easily transfer tools of financial mathematics to energy modelling. Firstly, the energy storage is either impossible or very expensive, mainly in the form of hydroelectric potential ([Aïd, 2015]). Therefore supply and demand must always match. Any disturbances of this equilibrium can result in large spikes in the electricity spot (day-ahead) market, which lead to a strong and heteroscedastic (time-varying) volatility of electricity prices, especially of spot contracts. A long-term underproduction of energy may result in a decreased frequency or even blackouts with dire consequences, such as the famous European blackout in 2006 ([European Commission, 2007]). To prevent such dangerous events, supply and demand must be adjusted in real time. Transmission system operators must also keep sufficient operating reserves, i.e. energy generators capable of producing electricity in a short time in case of emergency. We can already see the importance of accurate models of energy generation which help plan such reserves effectively.
Over time prices tend to a long-term average determined by the demand-supply balance, so they revert to the mean. Furthermore, electricity prices

exhibit seasonality: the demand significantly increases in winter months (due to the increased usage of heating and lights) as well as during hot summer months (air conditioning), although these patterns vary significantly depending on the local climate ([Aïd, 2015]). Because prices react to the demand, very cold or very warm weather usually results in more expensive electricity. The periodic behaviour occurs also at a smaller, weekly scale: the demand increases in the peak time, when people use electricity for their everyday activities. Finally, the distributions of prices have fat tails and long memory ([Aïd, 2015]).

2.3 Electricity markets

In this section we briefly introduce electricity markets; for a more detailed description we refer the reader to [Aïd, 2015]. Deregulated electricity markets exist since 1981, when Chile quoted the first electricity spot price. In 1993 the Scandinavian market introduced also futures contracts. Today liberalised energy markets exist all over the world, including Austria and Germany, the countries of interest in this part of the thesis. In 01.10.2018 this market split into two separate bidding zones ([European Energy Exchange, 2018]). However, since we consider only the data up to 2016, we introduce electricity markets using the example of the joint Austrian-German market. It consists of:

- The spot/day-ahead market (EPEX SPOT), where one can buy or sell energy delivered during the following day.
- The futures/forward market (EEX), where the participants trade electricity delivered later than one day ahead.
- The intraday/balancing market (EPEX SPOT), where the transmission system operator and the players adjust the traded quantities to ensure the balance between supply and demand.

In this thesis we focus on modelling the spot market, although in Chapter 4 we use the data from the futures market to calibrate our model.

2.3.1 The spot market

Every day registered spot market players bid to sell or buy electricity for delivery during chosen hours of the following day. If the delivery happens during hours of high demand (7 a.m. to 8 p.m.), we call the contract *peak load*; during low demand we talk about *off-peak* contracts. If the delivery period encompasses all hours, we call it *base load*. At noon the market is cleared for the following day, which means that the market organiser establishes the price for each delivery hour, based on all the bids.
Prices are constrained between -500 and 3000 EUR/MWh. Negative prices occur rarely, but we still observe them when the supply significantly exceeds the demand. Traditional coal power plants require a lot of time and resources to stop and start running, therefore sometimes energy providers prefer to pay the consumers for using electricity instead of switching the production off and on (for more details see [EPEX SPOT, 2019]). Negative prices directly result from our inability to store energy. Such events happen mostly at night when the demand tends to be lower than during the waking hours of the majority of people.

2.3.2 The forward market

Because all energy trading happens on exchanges, we use terms *forward* and *futures* market interchangeably. As opposed to the spot market, sellers cannot choose the hours of delivery, so they agree to deliver electricity during all hours of the delivery period: a year (called a *calendar*), a quarter, a month, a week, a weekend or a day. These delivery periods exist with different maturities, so in total the Austrian-German market offers the participants the following contracts: 6 calendars (1 year ahead up to 6 years ahead), 11 quarters, 9 months, 4 weeks, 2 weekends and 8 days. However, in our work we consider only six types of futures contracts: one month ahead (1MAH), two months ahead (2MAH), up to six months ahead (6MAH). Market participants can decide on either peak load, base load or off-peak contracts, although the latter option is available only for the yearly, quarterly and monthly futures.
Similarly to spot, futures prices exhibit seasonal patterns, in this case related

to the expiry date ([Borovkova and Geman, 2006]). Furthermore, the volatility of futures prices increases with the decreasing length of delivery period. Another important feature of futures prices is their relationship to spot prices. If futures contracts are cheaper than spot, we call this situation *backwardation*; the opposite event is called *contango*.

3

The data

3.1 Data description

In Chapter 4 and Chapter 5 we consider daily data from Austrian and German energy markets: electricity spot and monthly futures prices (see Section 2.3), total load (total energy generation) as well as wind energy generation. Unless stated otherwise, we estimate our models on five-year-long time series of daily data between 01.01.2011 and 31.12.2015. We downloaded hourly time series from the website of the Open Power System Data project ([Open Power System Data, 2019]) and Austrian Power Grid ([Austrian Power Grid, 2019]). We use the following variables.

- Austrian Power Grid (APG):
  - Day-ahead forecasts of wind energy production in Austria [MW].
  - Day-ahead forecasts of total load (total energy production) in Austria [MW].
- Open Power System Data:

- Day-ahead forecasts of wind energy production in Germany, summed over all four energy providers: TransnetBW, TenneT, 50Hertz and Amprion [MW].
- Actual total load (total wind energy production) in Germany [MW].
- Spot prices in the joint Austrian-German market [EUR].

These two data sources are not fully consistent with regard to the load as they provide values from different areas of Austria. The Open Power System Data gather data not only from the control area APG, but also from the rest of the country, including industrial production units and railroad consumption, not connected to the APG ([European Commission, 2018b]). However, in our research we are mostly interested in the impact of wind energy generation on energy prices. Since wind energy should not influence commercial production units, these inconsistencies do not cause problems in our analyses — especially as Austrian load is a few orders of magnitude smaller than German one, so any differences in the Austrian data become negligible when we look at both countries together. Figure 3.1 presents the control areas of four German transmission system operators.

Remark 3.1. The meaning of the term "load" varies across the literature: it can denote either power (the rate of electricity consumption over time) in kW or MW or consumed energy in kWh or MWh ([Hong, 2014]). Although hourly data sets will report the same magnitude regardless of the meaning, in our research we consider also other time periods, so for consistency we report the data in MWh and EUR/MWh. We recall that we define a watt as one joule per second, so it measures the power produced or used in a given time – in other words, the rate of energy generation or consumption. We also recall that

$$1\,\mathrm{kW} = 10^3\,\mathrm{W}$$
$$1\,\mathrm{MW} = 10^6\,\mathrm{W}$$
$$1\,\mathrm{GW} = 10^9\,\mathrm{W}.$$

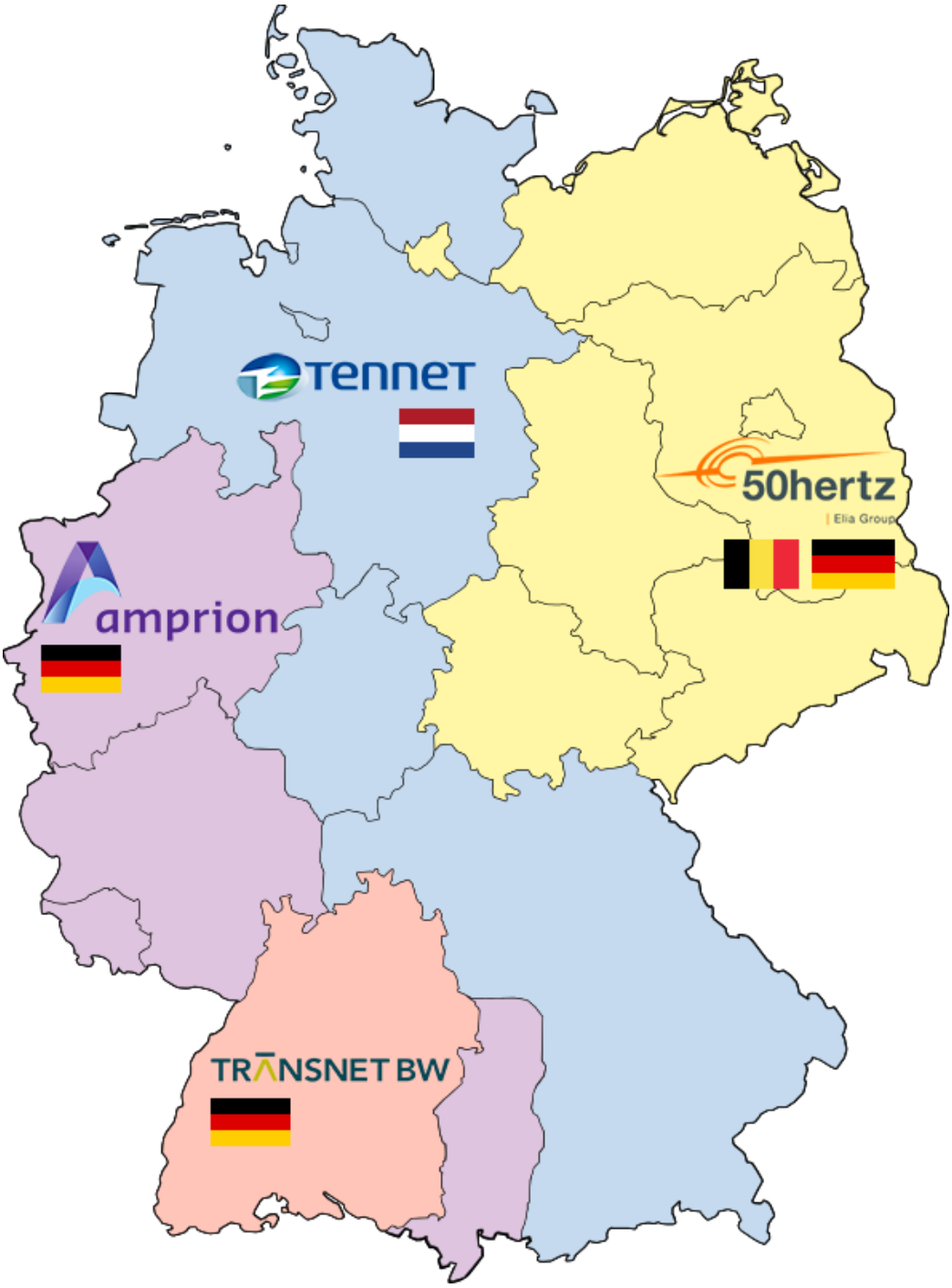


Figure 3.1: Control areas of transmission system operators in Germany ([McLloyd, 2013]).

On the other hand, watt-seconds are units of energy defined as one joule. For example, a country with hourly demand of 20 GW will use 20 GWh in one hour. A country with hourly demand of 10 GW will use the same amount of energy in two hours. Conversions of energy units are similar to power units:

$$1\,\mathrm{kWh} = 10^3\,\mathrm{W\,h}$$
$$1\,\mathrm{MWh} = 10^6\,\mathrm{W\,h}$$
$$1\,\mathrm{GWh} = 10^9\,\mathrm{W\,h}.$$

We aggregate the data to daily values: we average the prices and sum up total wind energy productions as well as total loads. Similarly to [Jónsson et al., 2010], we approximate the unavailable day-ahead forecasts of total load in Germany by randomly perturbing the actual total load. Precisely, we assume that for each day $t = 1, \ldots, N$ (N denotes the number of observed days)

$$\text{forecasted load}(t) = \text{actual load}(t) + \varepsilon(t),$$

where $\varepsilon \sim N(0, \sigma^2)$ and σ equals approximately 2% of averaged actual load. While this method works very well in practice ([Jónsson et al., 2010], [Veraart, 2016]), [Jónsson et al., 2010] point out some flaws of this approach. First, it cannot account for possibly autocorrelated residuals of forecasted load. Second, it assumes the independence of forecast errors of wind energy production and total load, which is unrealistic as they both stem from weather forecasts. However, the effect of these simplifications on our applications is negligible. Alternatively, we could obtain the forecasted load by smoothing the actual load, as we expect the forecasts to be smoother than the data.
We sum up the values of forecasted load and wind energy production in both countries to get figures for the whole region. Finally, we convert the units to gigawatts. In the rest of this chapter "wind" denotes the sum of day-ahead forecasts (true or perturbed) of wind energy production levels in Austria and Germany (in GWh).

3.2 Futures prices

In Chapter 4 we propose a model calibration procedure which additionally requires the prices of monthly futures contracts: one month ahead (1MAH) up to six months ahead (6MAH) [European Energy Exchange, 2017]. Figure 3.2 shows the empirical data for 1MAH and 6MAH, two most extreme time horizons considered. Because the trading periods of futures contracts are limited, we can see discontinuities in the beginning and the end of trading periods. We also note seasonal patterns in the data as well as the downward trend over time. Finally, the spikes characteristic to spot prices are not present in the futures prices, which [ter Haar, 2010] explains by the smaller influence of demand-supply imbalances on longer term contracts. Table 3.1 presents the summary statistics of considered futures data.
The main topic of this thesis is spot prices. Therefore in Table 3.1 we present the summary statistics of considered futures data and for more details about futures contracts, including a thorough data analysis, we refer the reader to [ter Haar, 2010] and references therein.

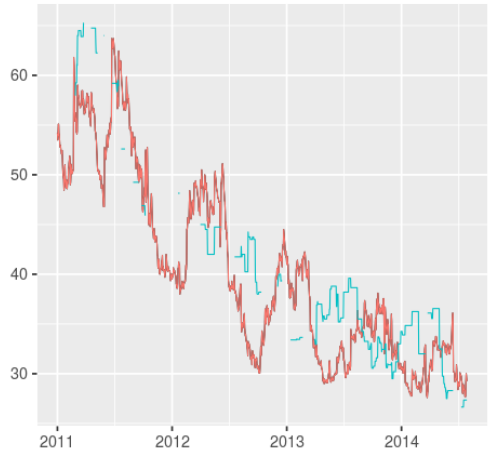


Figure 3.2: Time series of futures prices data: one-month-ahead (red) and six-months-ahead (blue).

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1MAH	27.60	32.45	38.45	40.33	48.16	63.70
2MAH	26.40	32.54	38.50	40.96	48.56	67.80
3MAH	26.40	33.60	38.58	41.35	47.50	67.50
4MAH	28.00	33.17	38.15	41.33	46.88	65.50
5MAH	24.40	33.58	38.65	40.85	47.00	66.00
6MAH	26.65	33.25	36.55	39.40	42.00	65.25

Table 3.1: Summary statistics of monthly futures prices computed after removing NA values.

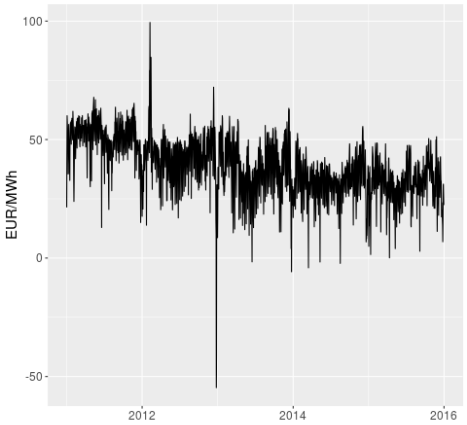
3.3 Exploratory data analysis

Table 3.2 presents the summary statistics of daily data sets: day-ahead electricity prices, day-ahead forecasts of wind energy production and total load, denoted by S, WD and LD, respectively. First we note the high variability of all time series. We would like to point out the occurrence of daily negative prices and the strict positivity of daily wind energy production forecasts. Figure 3.3a, Figure 3.4a and Figure 3.5a illustrate the same data with time series plots, in which we can clearly see the seasonal patterns. Furthermore, during the time of observations spot prices have significantly decreased while wind energy production increased. Also, Figure 3.3b, Figure 3.4b and Figure 3.5b, the plots of autocorrelation functions, indicate strong seasonal patterns in all variables.

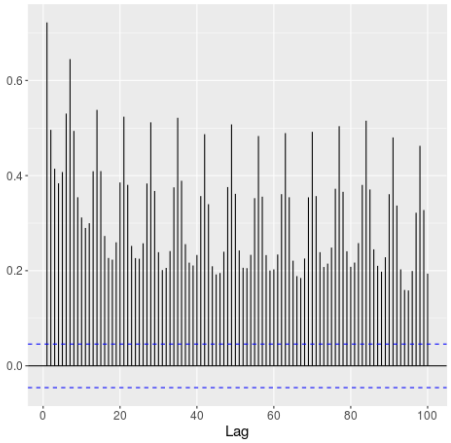
In Figure 3.6 we observe the densities as well as the relationships (scatter plots and correlation coefficients) between variables. We also see that wind energy production forecasts are clearly right-skewed. Finally, the correlation coefficients between S and WD as well as between S and LD indicate some level of dependency between these variables.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
S	-54.70	31.26	38.76	39.18	48.35	99.43
WD	2.21	67.44	117.93	156.30	204.42	792.91
LD	1103.53	1441.13	1617.32	1586.58	1727.01	2005.90

Table 3.2: Summary statistics of day-ahead electricity prices S [EUR/MWh], day-ahead forecasts of daily wind energy production WD [GWh] and day-ahead forecasts of daily load LD [GWh].

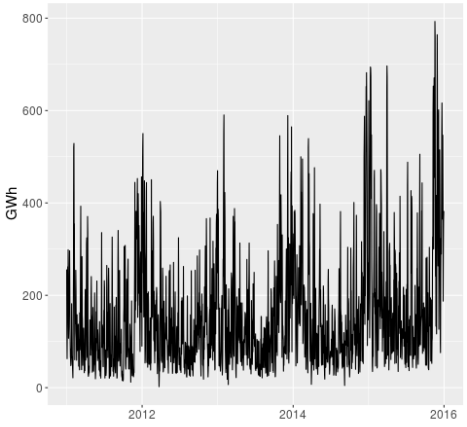


(a) Time series plot of electricity spot prices S.

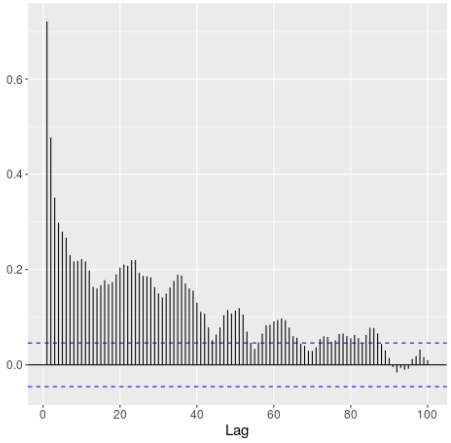


(b) Autocorrelation function of electricity spot prices S.

Figure 3.3: Time series and autocorrelation function plots of electricity spot prices S.

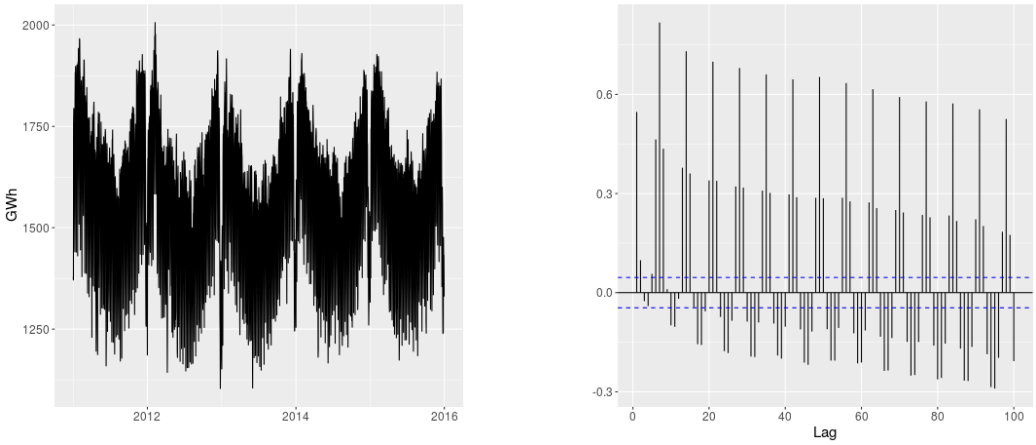


(a) Time series plot of day-ahead forecasts of daily wind energy production WD.



(b) Autocorrelation function of day-ahead forecasts of daily wind energy production WD.

Figure 3.4: Time series and autocorrelation function plots of day-ahead forecasts of daily wind energy production WD.



(a) Time series plot of day-ahead forecasts of daily load LD.

(b) Autocorrelation function of day-ahead forecasts of daily load LD.

Figure 3.5: Time series and autocorrelation function plots of day-ahead forecasts of daily load LD.

3.3.1 Wind penetration index and residual demand

Our research focuses on the relationship between electricity spot prices and wind energy production forecasts. We stress that it is the wind energy production forecasts, not the actual values, that impact the price, as they represent the information available to market participants at the time of transaction ([Ketterer, 2014]). It is crucial that we focus on day-ahead forecasts, because longer forecasting horizons reduce their accuracy and usefulness for making market decisions.

[Jónsson et al., 2010] argue that the same level of wind energy production will affect the price in a different way depending on the total demand. In the literature we find two ways of combining these two variables: wind penetration index, suggested for example by [Jónsson et al., 2010], as well as the residual load (or demand) studied by [Nicolosi and Fürsch, 2009] and others. Therefore we study additional time series: wind penetration index WPI, defined as the ratio between wind energy production and total load, as well as residual demand RD, defined as the difference between load and wind energy production. WPI is a dimensionless quantity with values between zero (no wind energy production) and one (all energy production coming

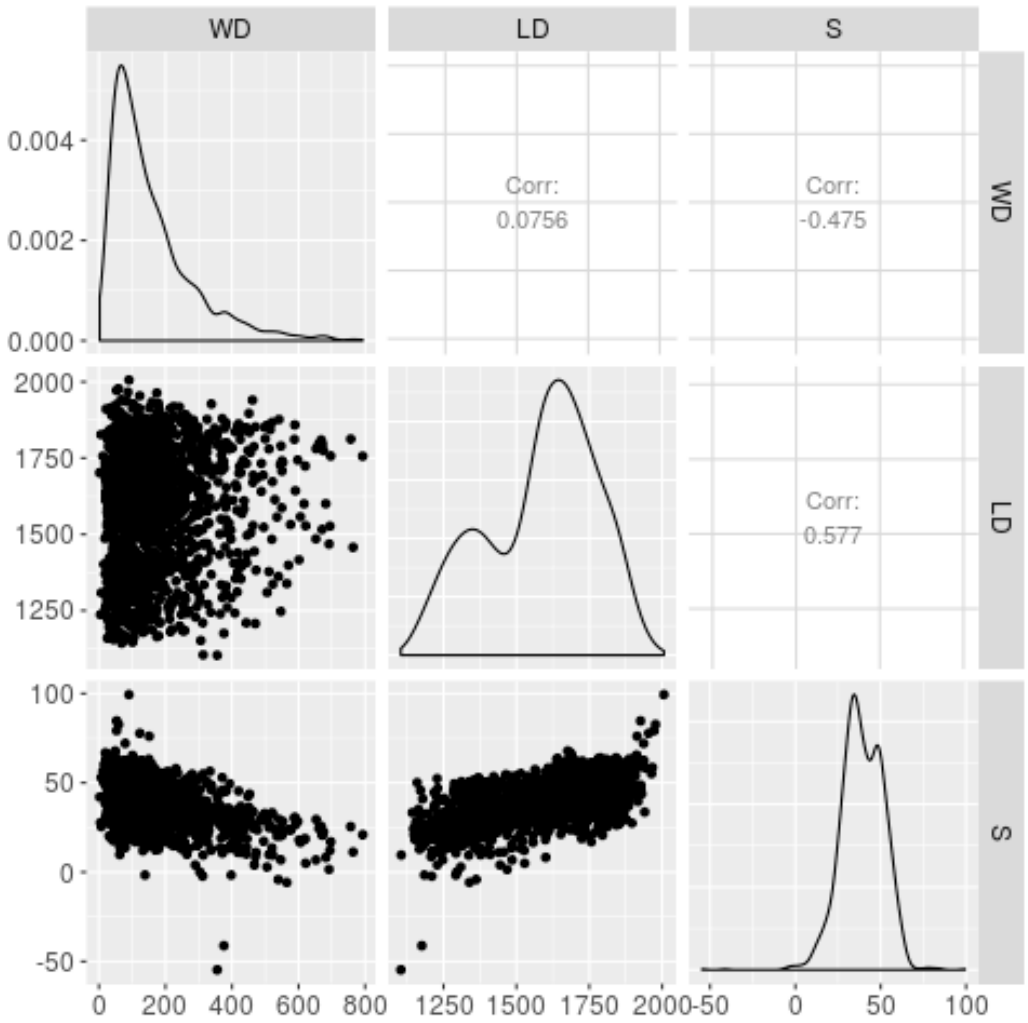
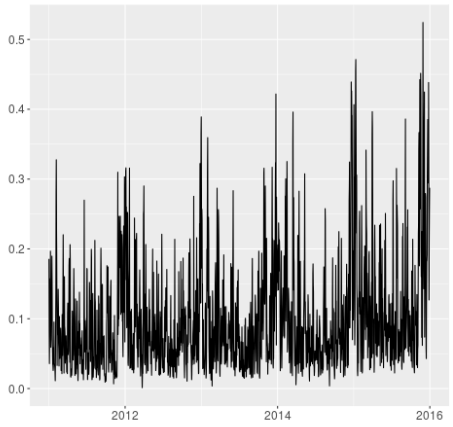


Figure 3.6: Relationships between day-ahead forecasts of wind energy generation WD and total load LD as well as electricity price S.

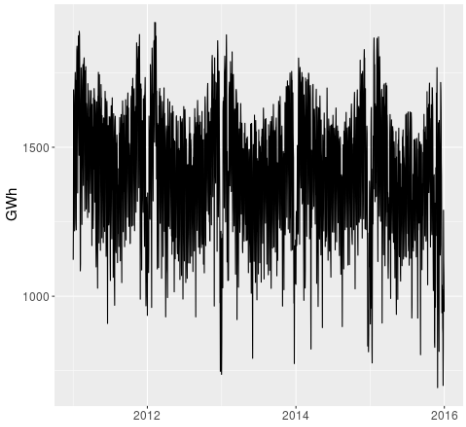
from wind energy). Because of energy export, on rare occasions the value of WPI might exceed one. RD corresponds to energy from sources other than wind, expressed in GWh. Table 3.3 shows the summary statistics of these two time series. In the discussed period (2011–2015) on average 10% of total energy produced in Austria and Germany came from wind. Figure 3.7a and Figure 3.7b present both time series. We note that WPI is very volatile and oscillates a lot around its mean level.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
WPI	0.00	0.04	0.07	0.10	0.13	0.52
RD	693.18	1267.06	1471.30	1430.28	1590.27	1919.02

Table 3.3: Summary statistics of day-ahead wind penetration index WPI [-] and day-ahead residual demand RD [GWh].



(a) Time series plot of day-ahead forecasts of the wind penetration index WPI.



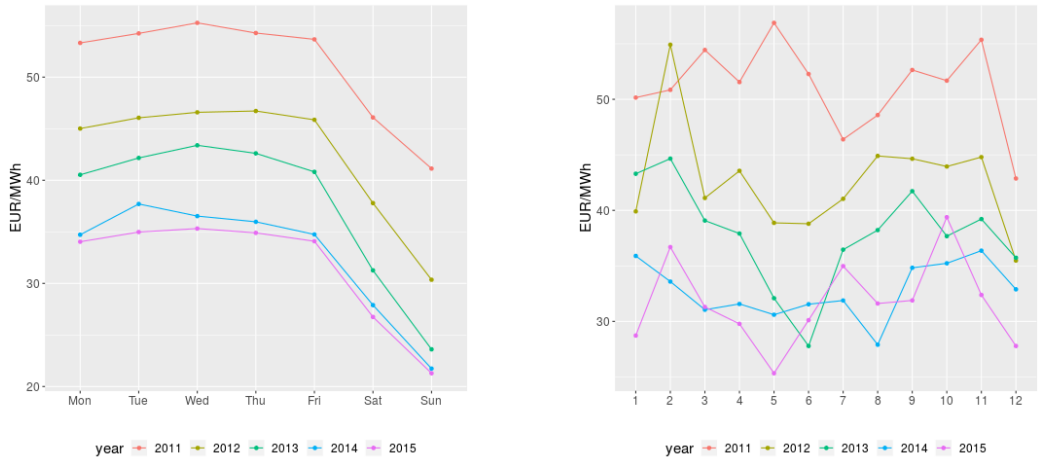
(b) Time series plot of day-ahead forecasts of the residual demand RD.

Figure 3.7: Time series and autocorrelation function plots of day-ahead forecasts of WPI and RD.

3.4 Seasonality

3.4.1 Seasonal patterns in the data

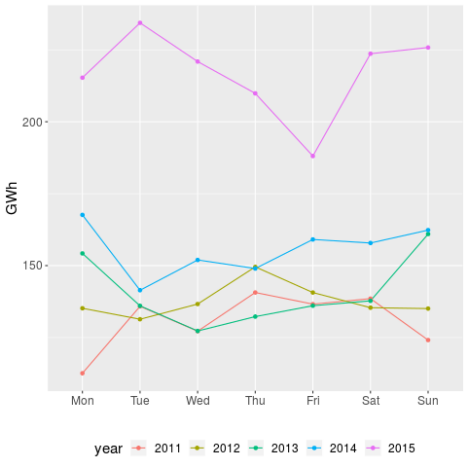
All variables considered exhibit seasonal patterns. According to Figure 3.8, on average lowest prices occur during the weekends and in summer months. We also see that the prices decreased from 2011 to 2015. Figure 3.9 shows that while (unsurprisingly) we generate similar amounts of wind energy regardless of the weekday, the production is much higher in winter than in summer. Moreover, we notice a significant increase in wind energy production as the years progressed, which might explain the decrease in prices. When it comes to the total load and residual demand, in Figure 3.10 and Figure 3.11 we can see that similarly to spot prices lowest values occur during the weekends as well as in summer. The total energy generation stayed approximately the same over the years. Finally, from Figure 3.12 we learn that the highest values of the wind penetration index occur on the weekends and in winter, which is mostly due to lower values of total load in the denominator.



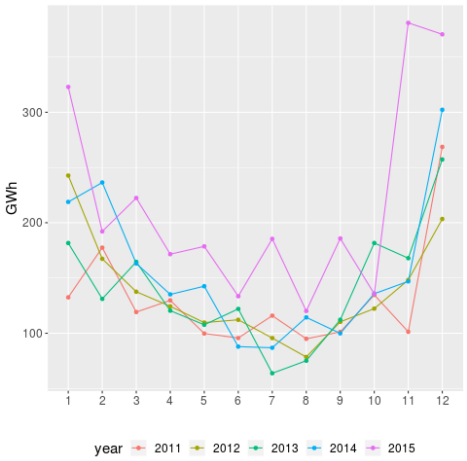
(a) Averaged day-ahead electricity prices S grouped by weekdays.

(b) Averaged day-ahead electricity prices S grouped by months.

Figure 3.8: Day-ahead electricity prices S grouped by seasons: weekdays (left panel) and months (right panel).

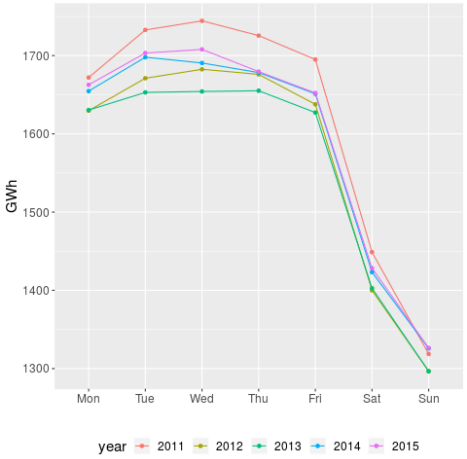


(a) Averaged day-ahead forecasts of daily wind energy production WD grouped by weekdays.



(b) Averaged day-ahead forecasts of daily wind energy production WD grouped by months.

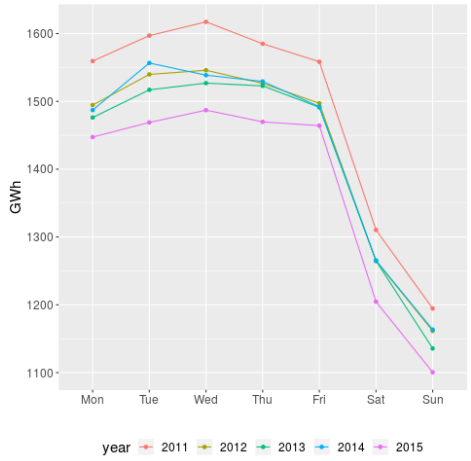
Figure 3.9: Day-ahead forecasts of daily wind energy production WD grouped by seasons: weekdays (left panel) and months (right panel).



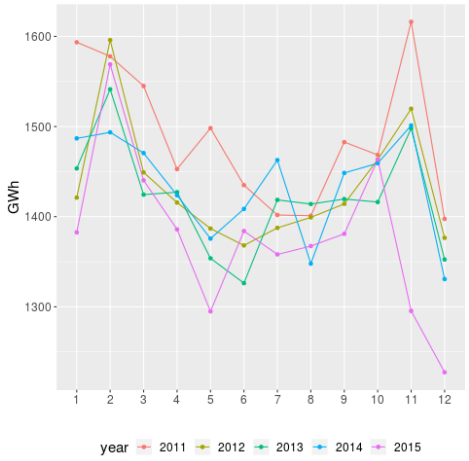
(a) Averaged day-ahead forecasts of daily load LD grouped by weekdays.

(b) Averaged day-ahead forecasts of daily load LD grouped by months.

Figure 3.10: Day-ahead forecasts of daily load LD grouped by seasons: weekdays (left panel) and months (right panel).

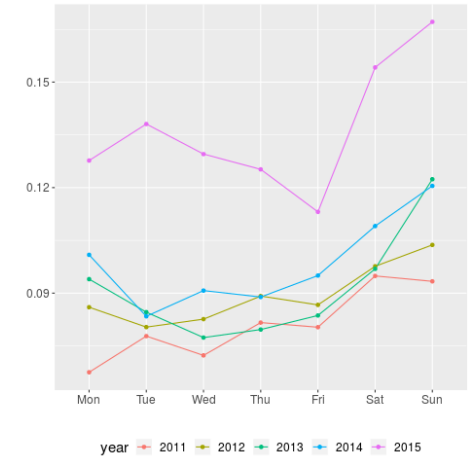


(a) Averaged day-ahead forecasts of the residual demand RD grouped by weekdays.

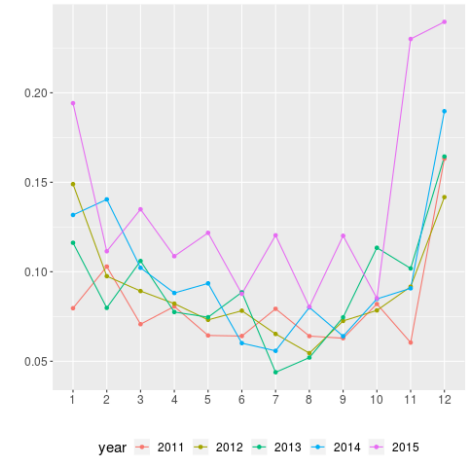


(b) Averaged day-ahead forecasts of the residual demand RD grouped by months.

Figure 3.11: Day-ahead forecasts of the residual demand RD grouped by seasons: weekdays (left panel) and months (right panel).



(a) Averaged day-ahead forecasts of the wind penetration index WPI grouped by weekdays.



(b) Averaged day-ahead forecasts of the wind penetration index WPI grouped by months.

Figure 3.12: Day-ahead forecasts of the wind penetration index WPI grouped by seasons: weekdays (left panel) and months (right panel).

3.4.2 Seasonality functions

Before we propose models of any of the variables, we remove trends and seasonal patterns from all data sets by assuming that

$$\begin{aligned} S(t) &= \Lambda^S(t) + \overline{S}(t), \\ WD(t) &= \Lambda^{WD}(t) + \overline{WD}(t), \\ LD(t) &= \Lambda^{LD}(t) + \overline{LD}(t), \\ WPI(t) &= \Lambda^{WPI}(t) + \overline{WPI}(t), \\ RD(t) &= \Lambda^{RD}(t) + \overline{RD}(t), \end{aligned}$$

where $\Lambda^\star$ denote deterministic seasonality and trend functions for appropriate variables, while bars detrended and deseasonalised variables. Let

$$\Lambda_1^\star(t) = c_0^\star + c_1^\star t + c_2^\star \cos\left(\frac{c_3^\star + 2\pi t}{365}\right),$$

where stars correspond to coefficients of the variable of interest. We additionally define dummy variables for days of the week ($d_{\text{Mon}}, \ldots, d_{\text{Sat}}$) and main holidays ($h(t) = 0$ for Dec 24th, Dec 25th, Dec 26th and Jan 1st; $h(t) = 1$ otherwise). We set

$$\Lambda_2^\star(t) = c_h^\star h(t) + c_{\text{Mon}}^\star d_{\text{Mon}}(t) + \cdots + c_{\text{Sat}}^\star d_{\text{Sat}}(t)$$

and

$$\Lambda_3^\star(t) = c_h^\star h(t) + c_{\text{Mon}}^\star d_{\text{Mon}}(t) + \cdots + c_{\text{Fri}}^\star d_{\text{Fri}}(t),$$

where the latter does not include a dummy variable for Saturday. Now we can define

$$\begin{aligned}
\Lambda^S(t) &= \Lambda_1^S(t) + \Lambda_2^S(t), \\
\Lambda^{WD}(t) &= \Lambda_1^{WD}(t), \\
\Lambda^{LD}(t) &= \Lambda_1^{LD}(t) + \Lambda_2^{LD}(t), \\
\Lambda^{WPI}(t) &= \Lambda_1^{WPI}(t) + \Lambda_3^{WPI}(t), \\
\Lambda^{RD}(t) &= \Lambda^{LD}(t) - \Lambda^{WD}(t).
\end{aligned}$$

As we do not observe weekly effects in WD, we remove only yearly seasonality. Furthermore, the dummy variable for Saturday is not significant in case of WPI, so we do not include this variable. Finally, since RD is an additive function of LD and WD, we do not introduce any additional seasonal factors for this variable.
We use the function `nls` from the `R` package `stats` ([R Core Team, 2018]) to estimate the coefficients of seasonality functions Λ. We initialise all parameters with 0.1. We present the estimated parameters in Table 3.4, Table 3.5, Table 3.6 and Table 3.7. The estimated coefficients reflect the patterns in the data. For example, c^S_{Sat} has the smallest magnitude of all weekly coefficients of Λ^S, which corresponds to lower spot prices on the weekends. Also, the trend coefficient c_1^{WD} is positive, which reflects the gradual increase in wind energy production over time.

3.4.3 Deseasonalied and detrended data

Figure 3.13, Figure 3.14, Figure 3.15, Figure 3.16 and Figure 3.17 present plots of deseasonalised time series averaged by weekdays or months. We can see that the deseasonalisation procedure was effective and removed most seasonal effects (please pay attention to the scale in the plots).
Since in this thesis we explore the relationship between electricity spot prices and wind energy production, we first need to check if such a relationship exists in deseasonalised data sets. In Figure 3.18 we observe strong correlations

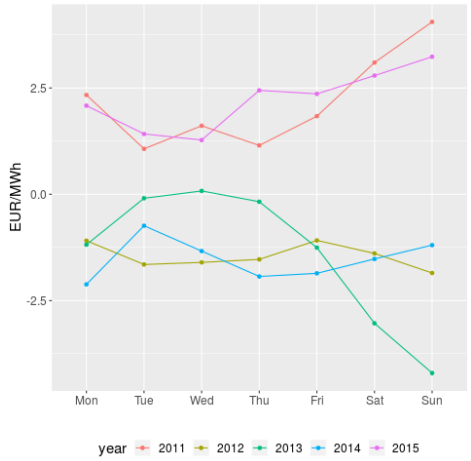
	Estimate	Std. Error	t value	$Pr(>\|t\|)$
c_0^S	9.757	1.9	5.135	3.119e-07
c_1^S	-0.0134	0.0003584	-37.4	2.038e-227
c_2^S	-3.027	0.2672	-11.33	8.465e-29
c_3^S	8328	32.17	258.9	0
c_{hol}^S	30.33	1.814	16.72	1.795e-58
c_{Mon}^S	13.92	0.698	19.94	3.931e-80
c_{Tue}^S	15.53	0.6981	22.25	3.38e-97
c_{Wed}^S	16.03	0.6982	22.96	1.303e-102
c_{Thu}^S	15.52	0.6982	22.22	5.643e-97
c_{Fri}^S	14.23	0.6987	20.37	3.359e-83
c_{Sat}^S	6.471	0.6981	9.27	5.122e-20

Table 3.4: Estimated parameters of the seasonality and trend function Λ^S.

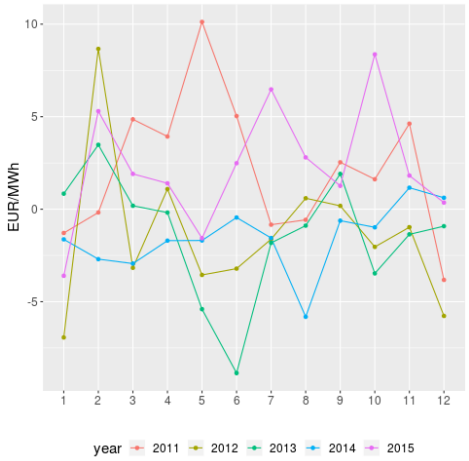
	Estimate	Std. Error	t value	$Pr(>\|t\|)$
c_0^{WD}	106.7	5.282	20.2	4.895e-82
c_1^{WD}	0.05425	0.005024	10.8	2.15e-26
c_2^{WD}	66.29	3.7	17.92	2.923e-66
c_3^{WD}	-4.364e+04	20.62	-2117	0

Table 3.5: Estimated parameters of the seasonality and trend function Λ^{WD}.

between $\overline{S}$ and three deseasonalised wind-related variables: $\overline{WD}$ (-0.600), $\overline{WPI}$ (-0.635) and $\overline{RD}$ (0.721). We also note that in first two cases the correlations are negative, while in the last one positive. This agrees with the expectations, since cheap wind energy tends to lower the prices (in case of the residual demand we subtract wind energy generation from the load, so we swap the sign).

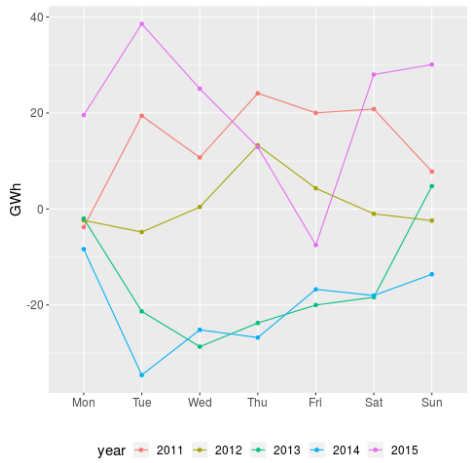


(a) Averaged deseasonalised day-ahead electricity prices $\overline{S}$ grouped by weekdays.

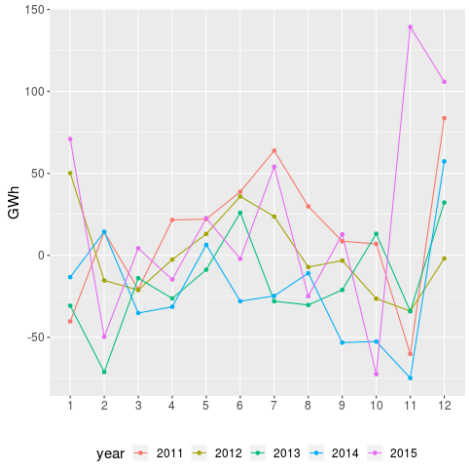


(b) Averaged deseasonalised day-ahead electricity prices $\overline{S}$ grouped by months.

Figure 3.13: Deseasonalised day-ahead electricity prices $\overline{S}$ grouped by seasons: weekdays (left panel) and months (right panel).

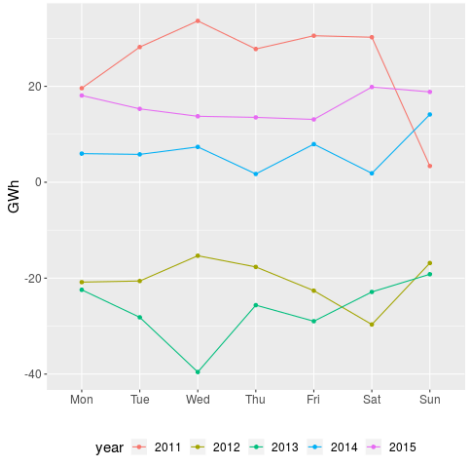


(a) Deseasonalised averaged day-ahead forecasts of daily wind energy production $\overline{WD}$ grouped by weekdays.

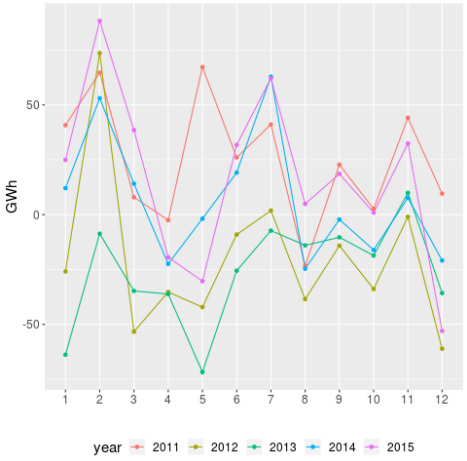


(b) Deseasonalised averaged day-ahead forecasts of daily wind energy production $\overline{WD}$ grouped by months.

Figure 3.14: Deseasonalised day-ahead forecasts of daily wind energy production $\overline{WD}$ grouped by seasons: weekdays (left panel) and months (right panel).

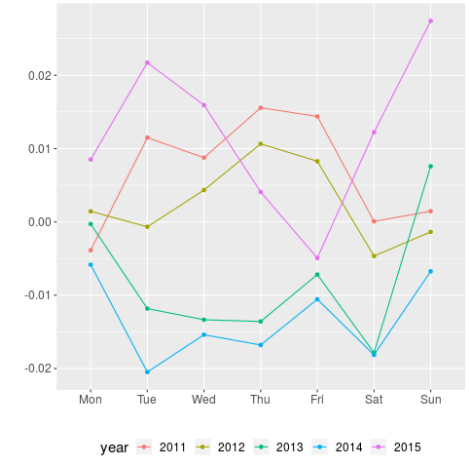


(a) Deseasonalised averaged day-ahead forecasts of daily load $\overline{LD}$ grouped by weekdays.

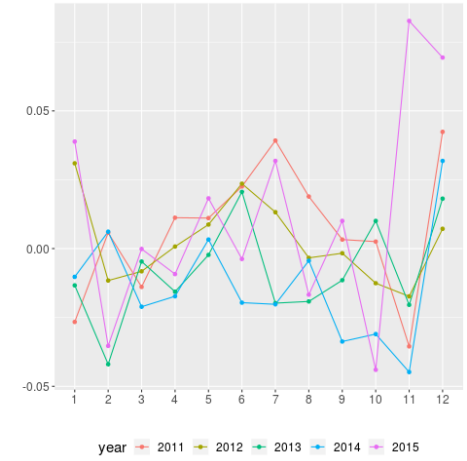


(b) Deseasonalised averaged day-ahead forecasts of daily load $\overline{LD}$ grouped by months.

Figure 3.15: Deseasonalised day-ahead forecasts of daily load $\overline{LD}$ grouped by seasons: weekdays (left panel) and months (right panel).



(a) Deseasonalised averaged day-ahead forecasts of the wind penetration index $\overline{WPI}$ grouped by weekdays.



(b) Deseasonalised averaged day-ahead forecasts of the wind penetration index $\overline{WPI}$ grouped by months.

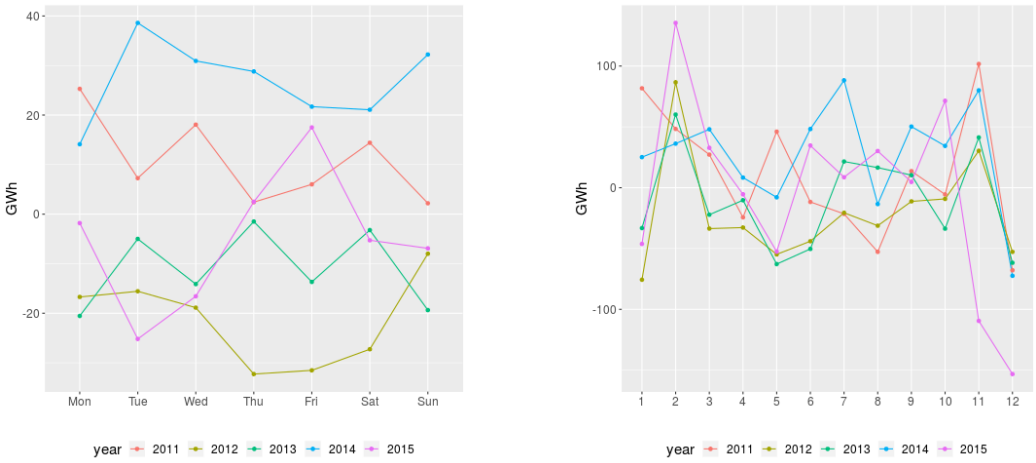
Figure 3.16: Deseasonalised day-ahead forecasts of the wind penetration index $\overline{WPI}$ grouped by seasons: weekdays (left panel) and months (right panel).

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
c_0^{LD}	868.2	20.77	41.81	4.178e-268
c_1^{LD}	-0.01136	0.003918	-2.899	0.003784
c_2^{LD}	-113	2.916	-38.74	8.935e-240
c_3^{LD}	-5.853e+04	9.441	-6199	0
c_{hol}^{LD}	458.2	19.83	23.11	8.758e-104
c_{Mon}^{LD}	337.1	7.63	44.18	4.403e-290
c_{Tue}^{LD}	380.5	7.631	49.87	0
c_{Wed}^{LD}	386.7	7.632	50.67	0
c_{Thu}^{LD}	373.7	7.632	48.97	0
c_{Fri}^{LD}	340.4	7.637	44.57	1.059e-293
c_{Sat}^{LD}	109.7	7.631	14.38	1.692e-44

Table 3.6: Estimated parameters of the seasonality and trend function Λ^{LD}.

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
c_0^{WPI}	0.2015	0.01653	12.19	6.527e-33
c_1^{WPI}	3.422e-05	3.178e-06	10.77	2.962e-26
c_2^{WPI}	0.03468	0.002365	14.67	4.102e-46
c_3^{WPI}	-75.88	24.94	-3.043	0.002378
c_{hol}^{WPI}	-0.1179	0.01608	-7.332	3.388e-13
c_{Mon}^{WPI}	-0.02055	0.00536	-3.834	0.0001301
c_{Tue}^{WPI}	-0.02341	0.00536	-4.367	1.331e-05
c_{Wed}^{WPI}	-0.0262	0.005361	-4.887	1.116e-06
c_{Thu}^{WPI}	-0.02371	0.005361	-4.423	1.032e-05
c_{Fri}^{WPI}	-0.02388	0.005367	-4.449	9.142e-06

Table 3.7: Estimated parameters of the seasonality and trend function Λ^{WPI}.



(a) Deseasonalised averaged day-ahead forecasts of the residual demand $\overline{RD}$ grouped by weekdays.

(b) Deseasonalised averaged day-ahead forecasts of the residual demand $\overline{RD}$ grouped by months.

Figure 3.17: Deseasonalised day-ahead forecasts of the residual demand $\overline{RD}$ grouped by seasons: weekdays (left panel) and months (right panel).

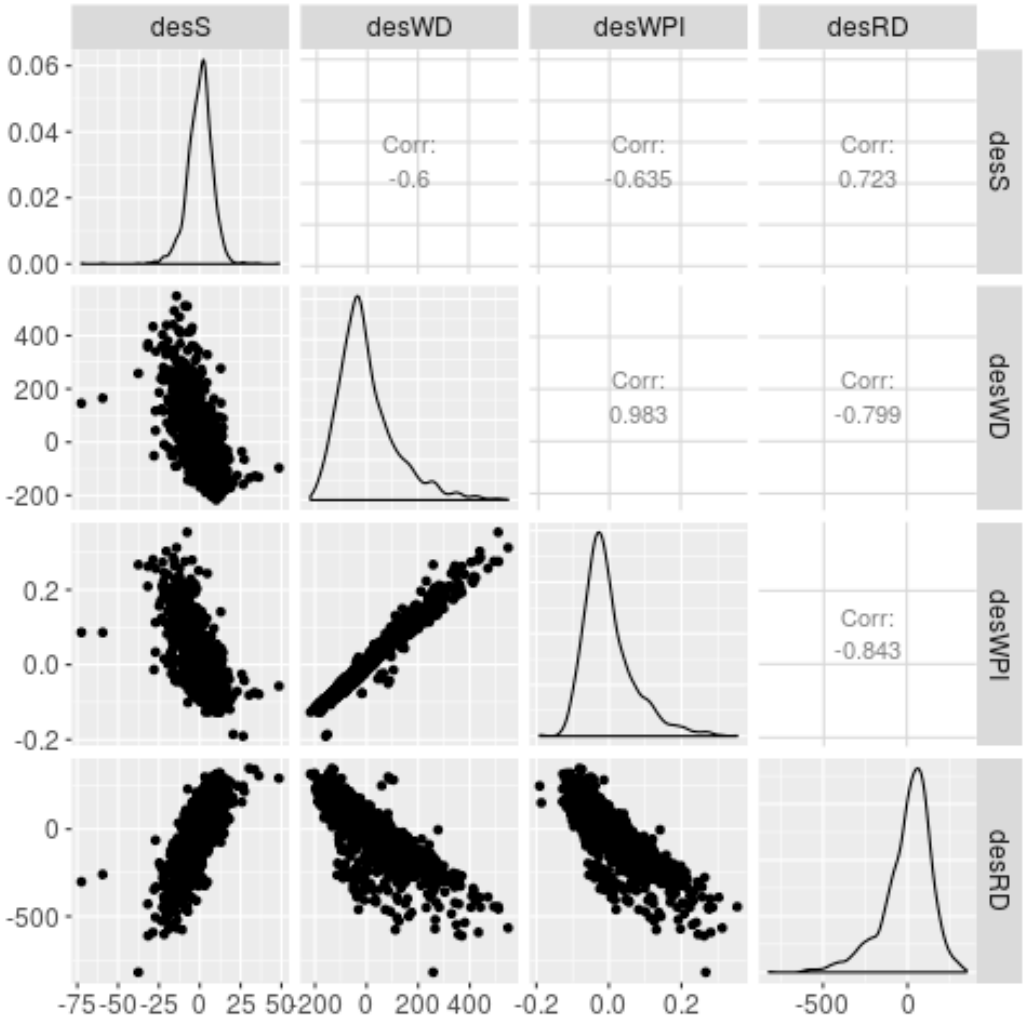


Figure 3.18: Relationships between deseasonalised electricity spot prices $\overline{S}$ and deseasonalised forecasts of wind energy production $\overline{WD}$, wind penetration index $\overline{WPI}$ and residual demand $\overline{RD}$.

4

A multifactor approach to modelling the impact of wind energy on electricity spot prices

In this chapter we introduce a model for electricity spot prices with wind energy production as an exogenous variable. Our work builds upon the paper by [Benth et al., 2014], who proposed an arithmetic model for spot prices with three factors: a deterministic seasonality and trend function as well as short- and long-term stochastic parts (details in Section 4.1).
To generalise this model we consider a short-term process more flexible than CARMA(2,1) proposed by [Benth et al., 2014], i.e. a Lévy semistationary (*LSS*) process ([Barndorff-Nielsen et al., 2013]) with stochastic volatility. Instead of α-stable driving process we choose processes from the generalised hyperbolic class (Appendix A) which not only capture the features of electricity prices but also are easy to fit thanks to the algorithm implemented by [Lüthi and Breymann, 2016].
After establishing the basic model, we extend it by adding wind-related exogenous variables. We show that the information about wind energy generation

improves the goodness of fit. Finally, we complete the research by modelling wind variables directly.
The chapter is structured as follows. In Section 4.1 we introduce a three-factor arithmetic model for spot prices and derive a formula for futures prices. In Section 4.2 we describe how to fit the model to empirical data and study all model terms in detail. We also present numerical results and discuss the possibility of adding stochastic volatility to the model (Subsection 4.2.3). Section 4.3 includes possible modifications of the basic model which include the information about wind energy generation. In that section we discuss the numerical results and compare the performance of different model modifications. In Section 4.4 we complete the work by modelling wind-related variables directly. Finally, in Section 4.5 we present the conclusions of our research.

4.1 The arithmetic model

The multifactor arithmetic model of spot prices introduced by [Benth et al., 2014] serves as a great starting point for inclusion of the information about wind energy production. It allows us to capture both short- and long-term price dynamics. Since the reliability of wind energy generation forecasts decreases rapidly with the forecasting horizon ([Met Office, 2019]), it is safe to assume that this additional information influences electricity spot prices only in short-term. Therefore the modelling framework suggested by [Benth et al., 2014] enables us to separate the dynamics of electricity prices into two parts, one of which depends on wind energy.

4.1.1 Assumptions

Let $S(t)$ denote the time series of electricity spot prices. Following [Benth et al., 2014], we propose the arithmetic model

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

where $\Lambda(t) + Z(t)$ form the long-term factor, while $Y(t)$ describes the short-term behaviour. Precisely, $\Lambda(t)$ denotes a deterministic seasonality and trend function, $Z(t)$ is a Lévy process with zero mean (under the physical measure) and

$$Y(t) = \int_{-\infty}^{t} g(t-s)\sigma(s-)dL(s)$$

with a deterministic kernel g such that $\lim_{t\to\infty} g(t) = 0$. Following [Barndorff-Nielsen et al., 2013] and references therein, we state conditions under which this integral exists.

Remark 4.1. We denote the Lévy triplet of $L(t)$ associated with a truncation function $h(z) = \mathbb{1}_{\{|z|\leq 1\}}$ by (d, b, l_L), where d denotes the drift, b the variance of the Gaussian component and l_L the Lévy measure. Let us define $\phi_t(s) := g(t-s)\sigma(s-)$. Then the process $(\phi_t(s))_{s\leq t}$ is integrable with respect to L if and only if $(\phi_t(s))_{s\leq t}$ is $\mathcal{F}$-predictable and these three conditions hold almost surely:

$$b\int_{-\infty}^{t} (\phi_t(s))^2 \, ds < \infty, \tag{4.1}$$

$$\int_{-\infty}^{t}\int_{-\infty}^{\infty} \left(1 \wedge |\phi_t(s)z|^2\right) l_L(dz)ds < \infty, \tag{4.2}$$

$$\int_{-\infty}^{t} \left| d\phi_t(s) + \int_{-\infty}^{\infty} \left(h\left(z\phi_t(s)\right) - \phi_t(s)h(z)\right) l_L(dz) \right| < \infty. \tag{4.3}$$

To ensure the square integrability, we assume that $L(t)$ has a finite second moment and replace Eq. (4.1) by

$$\int_{-\infty}^{t} \mathbb{E}\left[\phi_t(s)^2\right] ds = \int_{-\infty}^{t} g(t-s)^2 \mathbb{E}\left[\sigma(s)^2\right] ds < \infty$$

and

$$\mathbb{E}\left[\left(g(t-s)\sigma(s)ds\right)^2\right] < \infty.$$

For the latter condition it is enough to ensure that for some $a \in (0,1)$

$$\int_{0}^{\infty} g^{2a}(x)dx < \infty$$

and

$$\int_{-\infty}^{t} g^{2(1-a)}(t-s)\mathbb{E}\left[\sigma(s)^2\right].$$

The short-term process belongs to a class of Lévy semistationary (LSS) processes (cf. [Barndorff-Nielsen et al., 2013]), so $Y(t)$ is stationary if and only if $\sigma(t)$ and the increments of $L(t)$ are jointly stationary. It is worth noting that while integrating from $-\infty$ does not correspond to the real world, where $t \geq 0$, we need such an integral to obtain a stationary model. Therefore, similarly to [Barndorff-Nielsen et al., 2013], we assume that $s(0)$ is a realisation of a random variable $\Lambda(0) + Z(0) + Y(0)$. Here $\sigma(t)$ is a càdlàg stochastic process describing the volatility of $Y(t)$. While the empirical studies (Section 4.2) do not indicate much stochastic volatility in our data, we develop the theoretical framework which can accommodate $\sigma(t)$ with a short memory. We suggest using a (stationary) Ornstein-Uhlenbeck process, i.e. $\sigma(t) = \int_{-\infty}^{t} e^{-\delta(t-s)} dV(s)$ with a constant $\delta > 0$ and a Lévy subordinator $V(t)$, independent from the driving Lévy process $L(t)$ (cf. Subsection 4.2.3).

4.1.2 Change of measure

In traditional financial mathematics, one needs to define a probability measure $\mathbb{Q}$ equivalent to the physical measure $\mathbb{P}$ which transforms the discounted price dynamics into a (local) $\mathbb{Q}$-martingale. However, in the electricity markets not all the assets are tradable, so $\mathbb{Q}$ can denote any probability measure equivalent to $\mathbb{P}$. As suggested by [Barndorff-Nielsen et al., 2013], we define the generalised Esscher transform of $L(t)$ with a parameter $\theta(t)$, a Borel measurable function, via the Radon-Nikodym density process

$$\left.\frac{d\mathbb{Q}_L^{\theta}}{d\mathbb{P}}\right|_{\mathcal{F}_T} = \exp\left(\int_0^t \theta(s)dL(s) - \int_0^t \phi_L\left(\theta(s)\right) ds\right).$$

Here $\theta(\cdot)$ is a real-valued function integrable with respect to $L(t)$ and $\phi_L(\cdot) = \log\left(\mathbb{E}[\exp(xL(1))]\right)$ denotes the log-moment generating function of $L(1)$ (if it exists).

By analogy, we define $\mathbb{Q}_V^\eta$ by

$$\left.\frac{d\mathbb{Q}_V^\eta}{d\mathbb{P}}\right|_{\mathcal{F}_T} = \exp\left(\int_0^t \eta(s)dV(s) - \int_0^t \phi_V\left(\eta(s)\right)ds\right). \tag{4.4}$$

Then we get a new probability measure for $Y(t)$: $\mathbb{Q}_{\mathbb{Y}} := \mathbb{Q}_L \times \mathbb{Q}_V$. We define a measure change for $Z(t)$ in a similar way, i.e.

$$\left.\frac{d\mathbb{Q}_Z^\kappa}{d\mathbb{P}}\right|_{\mathcal{F}_T} = \exp\left(\int_0^t \kappa(s)dZ(s) - \int_0^t \phi_Z\left(\kappa(s)\right)ds\right). \tag{4.5}$$

Finally, we define the probability measure $\mathbb{Q} := \mathbb{Q}_Y \times \mathbb{Q}_Z$. If we choose $\theta(t), \eta(t)$ and $\kappa(t)$ to be constant, then the change of measure will preserve the desirable Lévy property.

4.1.3 Futures contracts

Assuming no arbitrage, we can express the price $f_t(T)$ at time $t \geq 0$ for a futures contract with maturity $T \geq t$ as

$$\begin{aligned} f_t(T) &= \mathbb{E}_{\mathbb{Q}}\left[S(T)|\mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}[\Lambda(T) + Z(T) + Y(T)|\mathcal{F}_t] \\ &= \Lambda(T) + Z(t) + (T-t)\mathbb{E}_{\mathbb{Q}}\left[Z(1)\right] \\ &+ \int_{-\infty}^t g(T-s)\sigma(s-)dL(s) + \mathbb{E}_{\mathbb{Q}}\left[L(1)\right]\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds. \end{aligned} \tag{4.6}$$

Due to its nonstorability, electricity is delivered over a time period rather than at one specific moment. Thus for all $0 \leq t \leq T_1 < T_2$ we define the price of a futures contract with a delivery period $[T_1, T_2]$ by

$$F_t(T_1, T_2) := \mathbb{E}_{\mathbb{Q}}\left[\frac{1}{T_2 - T_1}\int_{T_1}^{T_2} S(T)dT \middle| \mathcal{F}_t\right]. \tag{4.7}$$

If following [Benth et al., 2014] we define time to maturity as $u := \frac{1}{2}(T_1 + T_2) - t$, then Eq. (4.7) becomes

$$\begin{aligned} F_t(T_1, T_2) = &\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \Lambda(T) dT + Z(t) + u\mathbb{E}_{\mathbb{Q}}\left[Z(1)\right] \\ &+ \frac{1}{T_2 - T_1} \left(\int_{T_1}^{T_2} \int_{-\infty}^{t} g(T-s)\sigma(s-)\, dL(s)\, dT \right. \\ &\left. + \mathbb{E}_{\mathbb{Q}}\left[L(1)\right] \int_{T_1}^{T_2} \int_{t}^{T} g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]\, ds\, dT \right). \end{aligned} \tag{4.8}$$

Now we are ready to study the long-term behaviour of the futures prices, which will help us conduct the empirical work Section 4.2.

Proposition 4.0.1. *Assume that* $\delta > 0$, $\lim_{x\to\infty} \int_0^x g(y) e^{-\frac{\delta}{2}(x-y)} dy = 0$ *and* $\sigma(t)^2 = \int_{-\infty}^{t} e^{-\delta(t-x)} dV(x)$. *Then for* $\tau > 0$ *and fixed* $t > 0$,

$$\begin{aligned} &\lim_{T_1\to\infty} \frac{1}{\tau} \left(\int_{T_1}^{T_1+\tau} \int_{-\infty}^{t} g(T-s)\sigma(s-)\, dL(s)\, dT \right. \\ &\left. + \mathbb{E}_{\mathbb{Q}}\left[L(1)\right] \int_{T_1}^{T_1+\tau} \int_{t}^{T} g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]\, ds\, dT \right) = \mathbb{E}_{\mathbb{Q}}\left[L(1)\right] C, \end{aligned} \tag{4.9}$$

where

$$C := \mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right] \int_0^{\infty} g(y) dy \tag{4.10}$$

and the limit is in the L^2-*sense.*

To prove Proposition 4.0.1 we will use the following lemma.

Lemma 4.0.1. *If* $\tilde{H} := \lim_{T\to\infty} H(T)$ *exists in the* L^2-*sense, then for a fixed* $\tau > 0$ *the* L^2-*limit* $\lim_{T_1\to\infty} \frac{1}{\tau} \int_{T_1}^{T_1+\tau} H(T) dT$ *also exists and equals* $\tilde{H}$.

Proof of Lemma 4.0.1. Our assumption says that

$$\lim_{T\to\infty} \mathbb{E}_{\mathbb{Q}}\left[(H(T) - \tilde{H})^2 \right] = 0,$$

so that for all $\varepsilon > 0$ there exists $\tilde{T}$ such that for all $T > \tilde{T}$ we have $\mathbb{E}_{\mathbb{Q}}\left[\left(H(T) - \tilde{H}\right)^2\right] < \varepsilon$. Thus if $T_1 > \tilde{T}$, then

$$\mathbb{E}_{\mathbb{Q}}\left[\left(\frac{1}{\tau}\int_{T_1}^{T_1+\tau} H(T)dT - \tilde{H}\right)^2\right] \leq \max_{T_1 \leq T \leq T_1+\tau} \mathbb{E}_{\mathbb{Q}}\left[\left(H(T) - \tilde{H}\right)^2\right] < \varepsilon, \tag{4.11}$$

which was to be proven. □

Proof of Proposition 4.0.1. First we will show that $\lim_{T_1\to\infty} \frac{1}{\tau}\int_{T_1}^{T_1+\tau}\int_{-\infty}^{t} g(T-s)\sigma(s-)\,dL(s)\,dT = 0$ in the L^2-sense. As all considered functions are measurable and non-negative, we can use Tonelli's theorem ([Tao, 2011, p. 171]) to compute

$$\begin{aligned}&\lim_{T\to\infty} \mathbb{E}_{\mathbb{Q}}\left[\left(\int_{-\infty}^{t} g(T-s)\sigma(s-)dL(s)\right)^2\right] \\ &= \mathbb{E}_{\mathbb{Q}}\left[L(1)^2\right]\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)^2\right]\int_{-\infty}^{t}\lim_{T\to\infty} g(T-s)^2 ds = 0,\end{aligned} \tag{4.12}$$

which by Lemma 4.0.1 proves this statement.
Now we need to prove that

$$\lim_{T_1\to\infty}\int_{T_1}^{T_1+\tau}\int_{t}^{T} g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]\,ds\,dT = C$$

in the L^2-sense. By Lemma 4.0.1 it is enough to prove that

$$\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_{t}^{T} g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds - C\right)^2\right] = 0.$$

We observe that

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds - C\right)^2\right] \\
&= \mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds\right)^2\right] \\
&- 2C\mathbb{E}_{\mathbb{Q}}\left[\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds\right] + C^2.
\end{aligned} \tag{4.13}$$

Using Jensen's inequality and Tonelli's theorem, we can estimate

$$\begin{aligned}
&\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds\right)^2\right] \\
&\geq \lim_{T\to\infty}\left(\mathbb{E}_{\mathbb{Q}}\left[\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds\right]\right)^2 \\
&= \lim_{T\to\infty}\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]\right]ds\right)^2 \\
&= \left(\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]\lim_{T\to\infty}\int_0^{T-t} g(y)dy\right)^2 = C^2.
\end{aligned} \tag{4.14}$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_{-\infty}^s e^{-\delta(s-x)}dV(x)}\middle|\mathcal{F}_t\right] \\
&\leq \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_{-\infty}^t e^{-\delta(s-x)}dV(x)} + \sqrt{\int_t^s e^{-\delta(s-x)}dV(x)}\middle|\mathcal{F}_t\right] \\
&= \sqrt{\int_{-\infty}^t e^{-\delta(s-x)}dV(x)} + \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV(x)}\right] \\
&= \sigma(t)e^{-\frac{\delta}{2}(s-t)} + \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV(x)}\right],
\end{aligned} \tag{4.15}$$

where to non-negative processes we apply the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ as well as the identity $\int_{-\infty}^t e^{-\delta(s-x)}dV(x) = e^{-\delta(s-t)}\int_{-\infty}^t e^{-\delta(t-x)}dV(x) = \sigma(t)^2 e^{-\delta(s-t)}$. We remark that $\sigma(s)$ is stationary in mean, i.e. $\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)\right] =$

$\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]$ for all $s \in \mathbb{R}$. Furthermore,

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV(x)}\right] = \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_0^{s-t} e^{-\delta u}dV(u)}\right] \\
&\leq \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_0^{\infty} e^{-\delta u}dV(u)}\right] = \mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right],
\end{aligned} \tag{4.16}$$

as we integrate a non-negative function over a smaller domain. Therefore

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]\mathbb{E}_{\mathbb{Q}}\left[\sigma(u)|\mathcal{F}_t\right]\right] \leq \mathbb{E}_{\mathbb{Q}}\left[\sigma(0)^2\right] e^{-\frac{\delta}{2}(s-t)}e^{-\frac{\delta}{2}(u-t)} \\
&+ \mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]\left(e^{-\frac{\delta}{2}(u-t)}\mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV(x)}\right]\right. \\
&\left.+\ e^{-\frac{\delta}{2}(s-t)}\mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^u e^{-\delta(u-x)}dV(x)}\right]\right) \\
&+ \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV(x)}\right]\mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^u e^{-\delta(u-x)}dV(x)}\right] \\
&\leq \mathbb{E}_{\mathbb{Q}}\left[\sigma(0)^2\right] e^{-\frac{\delta}{2}(s-t)}e^{-\frac{\delta}{2}(u-t)} \\
&+ \mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]^2\left(e^{-\frac{\delta}{2}(s-t)} + e^{-\frac{\delta}{2}(u-t)} + 1\right),
\end{aligned} \tag{4.17}$$

where the first inequality follows from Eq. (4.15) and the second one from Eq. (4.16). This implies that

$$
\begin{aligned}
&\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds\right)^2\right]\\
&=\lim_{T\to\infty}\int_t^T\int_t^T g(T-s)g(T-u)\mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]\mathbb{E}_{\mathbb{Q}}\left[\sigma(u)|\mathcal{F}_t\right]\right]dsdu\\
&\leq \mathbb{E}_{\mathbb{Q}}\left[\sigma(0)^2\right]\left(\lim_{T\to\infty}\int_t^T g(T-s)e^{-\frac{\delta}{2}(s-t)}ds\right)^2\\
&+2\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]^2\lim_{T\to\infty}\left(\int_t^T g(T-s)ds\int_t^T g(T-s)e^{-\frac{\delta}{2}(s-t)}ds\right)\\
&+\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]^2\left(\lim_{T\to\infty}\int_t^T g(T-s)ds\right)^2\\
&=\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)^2\right]\left(\lim_{T\to\infty}e^{-\frac{\delta}{2}(T-t)}\int_0^{T-t} g(y)e^{\frac{\delta}{2}y}dy\right)^2\\
&+2\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]^2\lim_{T\to\infty}\int_0^{T-t} g(y)dy\lim_{T\to\infty}\left(e^{-\frac{\delta}{2}(T-t)}\int_0^{T-t} g(y)e^{\frac{\delta}{2}y}dy\right)\\
&+\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]^2\lim_{T\to\infty}\int_0^{T-t} g(y)dy=C^2,
\end{aligned}
\tag{4.18}
$$

where we use Tonelli's theorem and assumptions of Proposition 4.0.1. From Eq. (4.14) and Eq. (4.18) we deduce that

$$
\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds\right)^2\right]=C^2. \tag{4.19}
$$

Because by Tonelli's theorem

$$
\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]ds\right]=\mathbb{E}_{\mathbb{Q}}\left[\sigma(0)\right]\lim_{T\to\infty}\int_t^T g(T-s)ds=C,
$$

Eq. (4.13) implies that $\lim_{T_1,\to\infty}\int_{T_1}^{T_1+\tau}\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)|\mathcal{F}_t\right]\,ds\,dT=C$ in the L^2-sense. □

Proposition 4.0.1 allows us to conclude that in the long end, i.e. for $t \ll T_1$, we can approximate the deseasonalised futures price by

$$\begin{aligned} \tilde{F}_t(T_1, T_2) &:= F_t(T_1, T_2) - \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \Lambda(T) dT \\ &\approx Z(t) + u\mathbb{E}_{\mathbb{Q}}\left[Z(1)\right] + \mathbb{E}_{\mathbb{Q}}\left[L(1)\right] C. \end{aligned} \tag{4.20}$$

4.2 Empirical studies

We fit the proposed model to two sets of Austrian and German data: daily averages of hourly spot prices and monthly (one month ahead, 1MAH, up to six months ahead, 6MAH) base load futures prices, traded between 1 January 2011 and 3 August 2015 (cf. Section 3.1). In order to split spot prices into $S(t) = \Lambda(t) + Z(t) + Y(t)$, we modify the algorithm proposed by [Benth et al., 2014, pp. 398-9]. The idea behind this method is to use futures prices to split spot prices into the long and short-term factors. In Proposition 4.0.1 we showed that in the long end deseasonalised futures prices are influenced only by the long-term factor $Z(t)$, therefore the algorithm requires us to choose a threshold $u^\star$ such that deseasonalised prices of contracts with times to maturity $u \geq u^\star$ are described approximately by $Z(t)$ alone; we recall that we defined the time to maturity as $u = \frac{1}{2}(T_1 + T_2) - t$. We fix $u^\star$ to 16 days, because the results do not seem to be sensitive to the choice of a particular threshold. We proceed as follows.

1. We estimate $\Lambda(t)$ from spot prices and subtract from $S(t)$, as described in Section 3.4.
2. We filter out a realisation of $Z(t)$ (Subsection 4.2.1).
3. We model $Y(t) = S(t) - \Lambda(t) - Z(t)$ (Subsection 4.2.2) as a CARMA(2,1) process.

We note that [Müller and Seibert, 2019] propose an alternative estimation procedure based on the Markov chain Monte Carlo method.

4.2.1 Non-stationary long-term factor $Z(t)$

Let us denote the empirical mean (i.e. the averaged data) by $\widehat{\mathbb{E}}[\cdot]$. For $u \geq u^\star$, using Proposition 4.0.1 and $\mathbb{E}\left[Z(t)\right] = 0$, we can approximate

$$\mu_{\tilde{F}}(u) := \widehat{\mathbb{E}}\left[\tilde{F}_t(T_1, T_2)\right] = u\widehat{\mathbb{E}}_{\mathbb{Q}}\left[Z(1)\right] + \widehat{\mathbb{E}}_{\mathbb{Q}}\left[L_1\right] C \tag{4.21}$$

and by linear regression estimate $\widehat{\mathbb{E}}_{\mathbb{Q}}\left[L_1\right] C$ and $\widehat{\mathbb{E}}_{\mathbb{Q}}\left[Z(1)\right]$, as presented in Figure 4.1. These parameters together allow us to recover a realisation of the

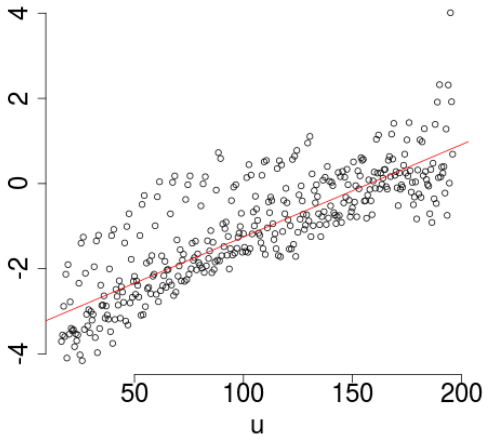


Figure 4.1: Plot of $\mu_{\tilde{F}}(u)$ with the fitted regression line.

long-term process

$$\begin{aligned}\widehat{Z}(t) &= \widehat{Z}\left(\frac{1}{2}(T_1 + T_2) - u\right) \\ &= \frac{1}{\mathrm{card}U(t, u^\star)} \sum_{(u,T_1,T_2)\in U(t,u^\star)} \left[\tilde{F}_t(T_1, T_2) - \widehat{\mathbb{E}}_{\mathbb{Q}}\left[L_1\right]\widehat{C} - u\widehat{\mathbb{E}}_{\mathbb{Q}}\left[Z(1)\right]\right],\end{aligned} \tag{4.22}$$

where

$$U(t, u^\star) := \{(u, T_1, T_2) \in \mathbb{R}^3 : u \geq u^\star \text{and} \quad \exists F_t(T_1, T_2) : \frac{1}{2}(T_1 + T_2) - t = u\}. \tag{4.23}$$

Because futures contracts are traded only from Monday to Friday, $Z(t)$ does

Figure 4.2: Estimated $Z(t)$.

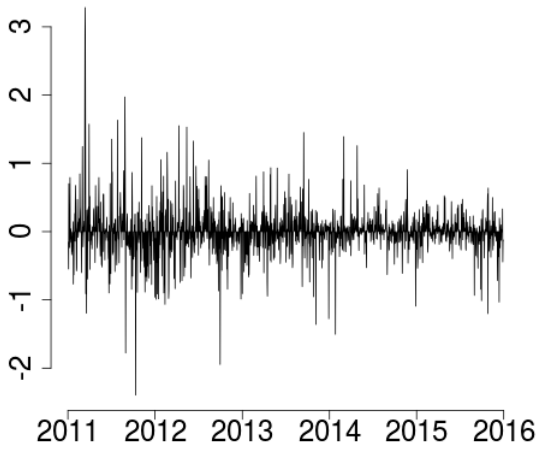


Figure 4.3: The increments of $Z(t)$.

not include the weekend data. Following [Benth et al., 2014], on the weekends we set this process to be constant and equal to the Friday price. The process $Z(t)$ is clearly non-stationary, therefore we study its increments, plotted in Figure 4.3. At first glance $Z(t)$ has stationary increments, as confirmed at 0.01 significance level by the augmented Dickey-Fuller (ADF) test, computed using the function `adf.test` from the `R` package `tseries` ([Trapletti and Hornik, 2018]). We observe that the estimated $Z(t)$ has (mostly) uncorrelated increments and squared increments, as shown in Figure 4.4 and Figure 4.5. Thus we can model $Z(t)$ by a Lévy process with increments following a suitable infinitely divisible distribution. Inspired by [Barndorff-Nielsen et al., 2013], we fit 11 distributions from a class of generalised hyperbolic distributions (Appendix A), using the `R` package `ghyp` provided by [Lüthi and Breymann, 2016]. We rank distributions according to the Akaike information criterion (AIC) in the ascending order, so that the first one gives the best fit. In Table 4.1 we can see that the increments of $Z(t)$ are best described by a symmetric normal inverse Gaussian (NIG) distribution. The fit is very good, as shown in Figure 4.6.

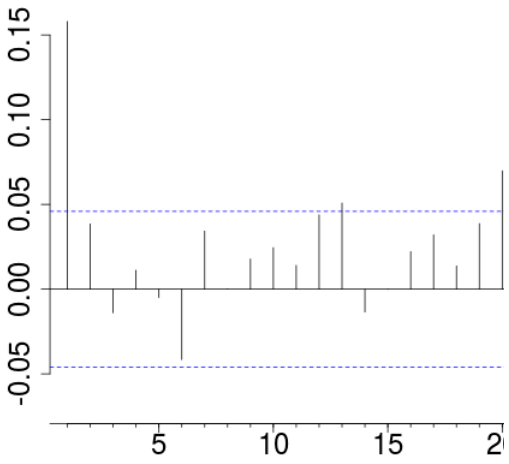


Figure 4.4: The autocorrelation function of the increments of $Z(t)$.

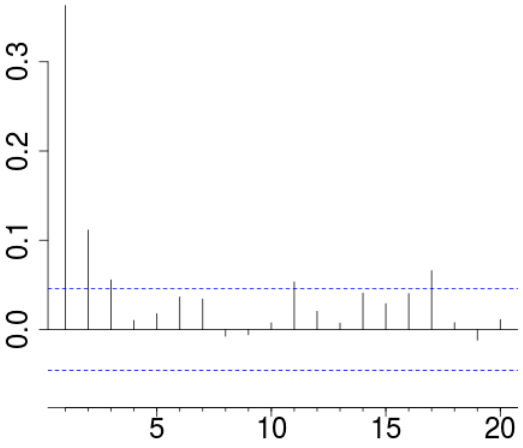


Figure 4.5: The autocorrelation function of the squared increments of $Z(t)$.

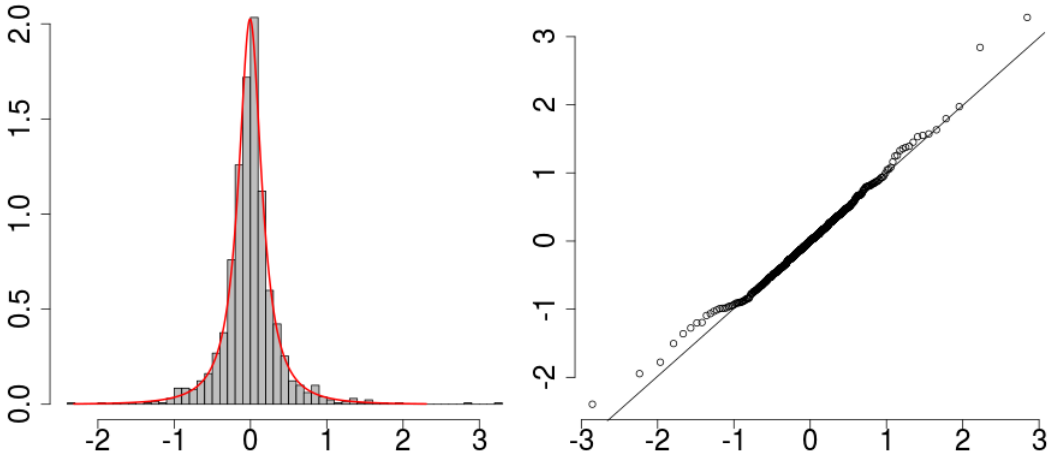


Figure 4.6: The increments of $Z(t)$ and the fitted generalised hyperbolic distribution: histogram and q-q plot.

	model	symmetric	lambda	alpha.bar	mu	sigma	gamma	aic	llh
8	NIG	TRUE	-0.500	0.236	-0.005	0.386	0.000	663.866	-328.933
6	ghyp	TRUE	-0.288	0.249	-0.005	0.381	0.000	664.904	-328.452
3	NIG	FALSE	-0.500	0.237	-0.008	0.385	0.007	665.605	-328.802
1	ghyp	FALSE	-0.294	0.250	-0.007	0.381	0.006	666.693	-328.346
10	t	TRUE	-1.036	0.000	-0.006	0.992	0.000	678.914	-336.457
5	t	FALSE	-1.039	0.000	-0.008	0.964	0.034	680.663	-336.331
4	VG	FALSE	0.636	0.000	0.022	0.347	-0.022	684.218	-338.109
9	VG	TRUE	0.727	0.000	0.002	0.356	0.000	697.183	-345.592
7	hyp	TRUE	1.000	0.027	-0.001	0.341	0.000	713.650	-353.825
2	hyp	FALSE	1.000	0.027	-0.001	0.341	0.001	715.649	-353.824
11	gauss	TRUE		Inf	-0.001	0.382	0.000	1194.509	-595.255

Table 4.1: Generalised hyperbolic distributions fitted to the increments of $Z(t)$ with parametrisation $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$.

4.2.2 Stationary short-term factor $Y(t)$

We obtain a realisation of $Y(t)$ by subtracting $Z(t)$ from the deseasonalised spot prices. Figure 4.7 shows the resulting process and its autocorrelation function. The augmented Dickey-Fuller (ADF) test ([Trapletti and Hornik, 2018]) allows us to reject the unit root hypothesis at significance level 0.01, so $Y(t)$ is a stationary time series.

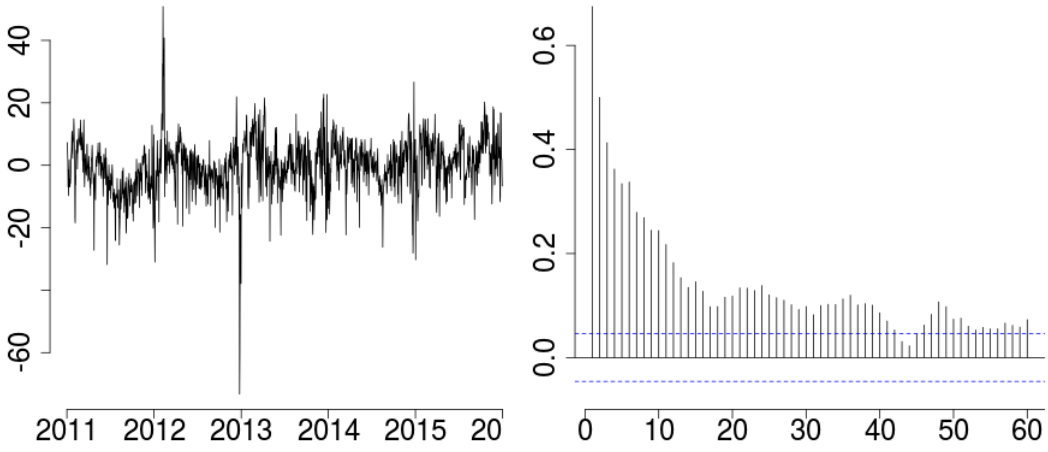


Figure 4.7: The estimated $Y(t)$: time series plot and the autocorrelation function.

Inspired by [Benth et al., 2014], we start by modelling $Y(t)$ with a CARMA(2,1) process, a special case of an Lévy semistationary processes described in Appendix C. The CARMA(2,1) kernel satisfies the assumptions of Proposition 4.0.1, as long as $\lambda_i \neq -\frac{\delta}{2}$ for $i = 1, 2$. This follows from

$$\lim_{x\to\infty} \int_0^x g(y) e^{-\frac{\delta}{2}(x-y)} dy = \lim_{x\to\infty} \sum_{i=1}^{2} \alpha_i \int_0^x e^{\lambda_i y} e^{-\frac{\delta}{2}(x-y)} dy =$$

$$\lim_{x\to\infty} \sum_{i=1}^{2} \frac{\alpha_i}{\lambda_i + \frac{\delta}{2}} \left(e^{\lambda_i x} - e^{-\frac{\delta}{2}x} \right) = 0,$$

because $\delta > 0$ and $\lambda_i < 0$ for $i = 1, 2$. We implement the algorithm from Section C.4 to fit a CARMA(2,1) model to $Y(t)$. First, we use the function `arima` from the `R` package `stats` ([R Core Team, 2018]) to fit an ARMA(2,1) process. We estimate the ARMA(2,1) model with parameters $(\phi_1, \phi_2, \theta) = (1.413, -0.446, -0.826)$. In Figure 4.8 we can see that the estimated discrete model describes $Y(t)$ accurately, although it cannot capture some extreme values.

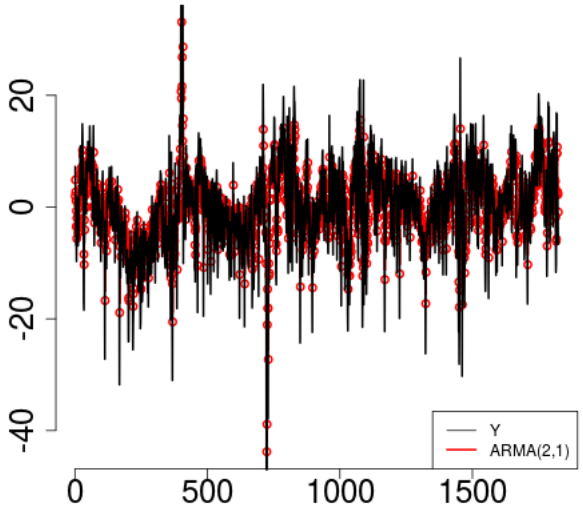


Figure 4.8: The ARMA(2,1) model fitted to $Y(t)$.

Proceeding as described in Section C.4, we map the estimated ARMA(2,1) parameters into CARMA(2,1) kernel parameters, which gives us the process

$$(D^2 + 0.809D + 0.048)Y(t) = (0.194 + D)DL(t),$$

where D denotes "differentiation" with respect to t (for details see Appendix C). We recover the driving Lévy noise $L(t)$. As with the increments of $Z(t)$, we choose the best fit to the increments ΔL of the driving noise among 11 generalised hyperbolic distributions. From Table 4.2 we learn that the best fit is given by the asymmetric Student's-t distribution. Below we describe the quality of these estimates, from now on called "true" estimates.

	model	symmetric	lambda	alpha.bar	mu	sigma	gamma	aic	llh
5	t	FALSE	-2.131	0.000	1.178	7.708	-1.215	12403.016	-6197.508
1	ghyp	FALSE	-2.055	0.319	1.186	7.654	-1.191	12405.297	-6197.649
10	t	TRUE	-2.044	0.000	0.293	7.854	0.000	12412.389	-6203.194
3	NIG	FALSE	-0.500	0.962	1.310	7.606	-1.342	12413.597	-6202.799
6	ghyp	TRUE	-2.047	0.005	0.293	7.850	0.000	12414.389	-6203.194
8	NIG	TRUE	-0.500	0.908	0.306	7.724	0.000	12425.469	-6209.734
2	hyp	FALSE	1.000	0.770	1.463	7.500	-1.498	12429.264	-6210.632
4	VG	FALSE	1.528	0.000	1.444	7.517	-1.479	12438.049	-6215.025
7	hyp	TRUE	1.000	0.724	0.286	7.604	0.000	12442.264	-6218.132
9	VG	TRUE	1.503	0.000	0.285	7.614	0.000	12451.107	-6222.553
11	gauss	TRUE		Inf	-0.034	7.949	0.000	12741.822	-6368.911

Table 4.2: Generalised hyperbolic distributions fitted to the increments ΔL of the driving noise of $Y(t)$.

To test the quality of the estimation procedure described in Section C.4, we simulate 1,000 paths of sampled the $Y(t)$ with $N = 1,824$ observations. As described in Section C.5, the simulation method requires simulating two Ornstein-Uhlenbeck processes driven by the same Lévy noise. We use the Euler-Maruyama method with the step size $h = 0.01$. We first simulate two Ornstein-Uhlenbeck processes of length $\frac{2N}{h} = 364,800$. We discard the first half of each time series as a burn-in period. Finally, we resample every $\frac{1}{h} = 100$ points to obtain two time series of length $N = 1,824$.
For each path we re-estimate the parameters par^{true}, which after averaging over all paths give us Monte Carlo estimates par^{MC}. We compute the bias

and standard error of all parameters:

$$\text{par}^{\text{MC}} - \text{par}^{\text{true}}$$

and

$$\sqrt{\frac{\sum_{i=1}^{1000} \left(\text{par}_i^{\text{MC}} - \text{par}^{\text{true}}\right)^2}{1000}},$$

respectively. We additionally report the relative bias and error defined as

$$\frac{\text{par}^{\text{MC}} - \text{par}^{\text{true}}}{\text{par}^{\text{MC}}}$$

and

$$\frac{\sqrt{\frac{\sum_{i=1}^{1000} \left(\text{par}_i^{\text{MC}} - \text{par}^{\text{true}}\right)^2}{1000}}}{\text{par}^{\text{MC}}},$$

respectively. From Table 4.3 we learn that the (relative) biases as well as (relative) standard errors of estimates are small. We present histograms of resulting estimates in Figure 4.9.

	a_1	a_2	b_0
True estimate	0.809	0.048	0.194
Monte Carlo estimate	0.811	0.045	0.191
Standard error	0.116	0.023	0.066
Relative error	0.143	0.519	0.348
Bias	0.002	-0.003	-0.004
Relative bias	0.002	-0.059	-0.019

Table 4.3: Parameters of CARMA(2,1) kernel estimated via bootstrapping with $1,000$ simulations.

Remark 4.2. It is crucial to simulate the Ornstein-Uhlenbeck processes on a sufficiently refined grid. For example, the Euler-Maruyama step size $ht = 1.0$ results in large (relative) biases of estimated parameters.

Similarly we re-estimate the parameters of the generalised hyperbolic noise. In Table 4.4 we observe large standard and relative errors as well as (relative) biases. However, we recall that to each path of the noise's increments we fit 11 distributions from the generalised hyperbolic class and choose the one

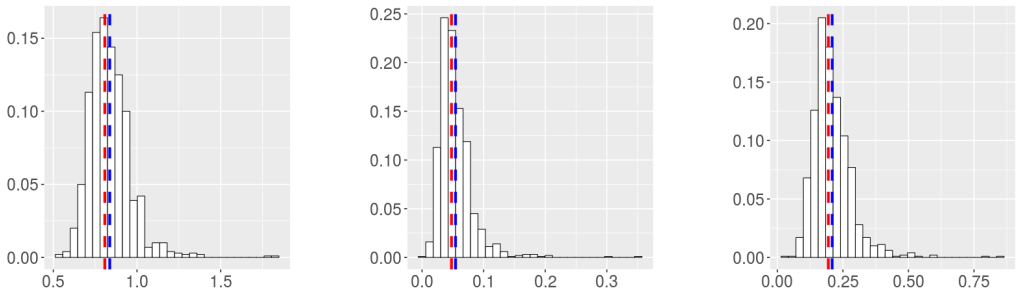


Figure 4.9: Parameters of CARMA(2,1) kernel estimated from $1,000$ simulated paths (from left to right: a_1, a_2 and b_0). Red vertical lines indicate true estimates, while red lines Monte Carlo estimates.

with the smallest value of AIC (Akaike information criterion). As stated in Section A.2, slightly different parameters might result in a distribution belonging to a different class. Therefore the distributions fitted to similar paths might have very different parameters. Figure 4.10 helps explain this phenomenon. For example, the estimates of λ and γ are clearly bimodal, so the average of parameters resulting from both cases matches neither of two distributions. However, this should not worry us because these different parameters result in very similar distributions.

	λ	α	δ	β	μ
True estimate	-2.131	0.020	11.592	-0.020	1.178
Monte Carlo estimate	0.631	0.415	29.214	-0.153	5.802
Standard error	38.008	1.212	20.699	1.062	9.926
Relative error	60.282	2.923	0.709	-6.944	1.711
Bias	2.761	0.394	17.622	-0.132	4.624
Relative bias	4.379	0.951	0.603	0.866	0.797

Table 4.4: Parameters of the generalised hyperbolic noise estimated via bootstrapping with $1,000$ simulations.

4.2.3 Stochastic volatility

Figure 4.11 and Figure 4.12 present the autocorrelation functions of the increments and squared increments of the recovered driving Lévy process $L(t)$. Figure 4.12 indicates that the squared increments of $L(t)$ are not

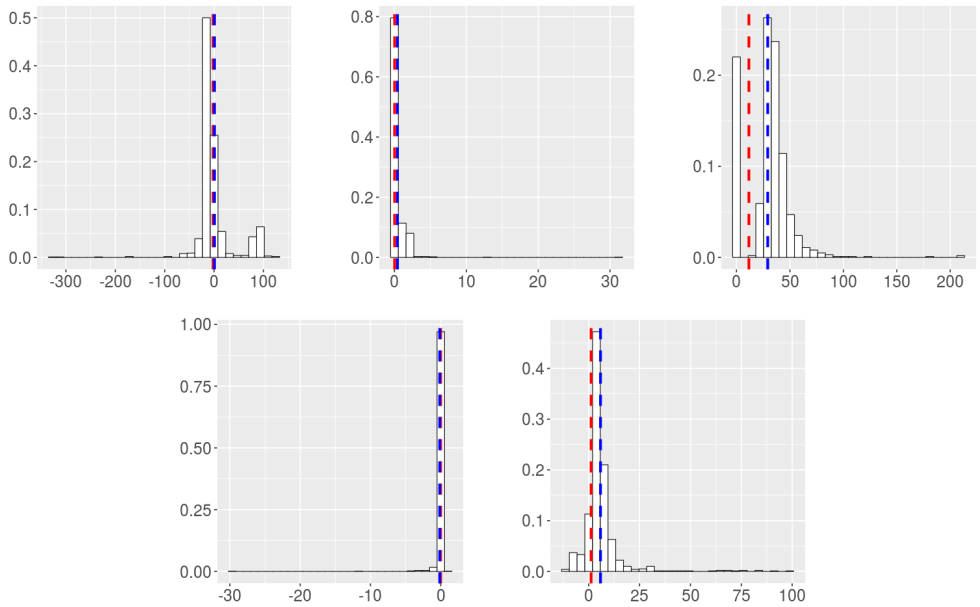


Figure 4.10: Parameters of the generalised hyperbolic noise estimated from $1,000$ simulated paths (from left to right: λ, α, δ, β and μ). Red vertical lines indicate true estimates, while blue lines Monte Carlo estimates.

independent, so the model could benefit from including a short-memory stochastic volatility in the definition of $Y(t)$. We notice the highest peak at lag 7, which could indicate that some traces of weekly behaviour remained after the deseasonalisation.

In this section again we assume that

$$Y(t) = \int_{-\infty}^{t} g(t-s)\sigma(s)dB(s),$$

where

$$\sigma(t)^2 = \int_{-\infty}^{t} e^{-\delta(t-s)}dV(s),$$

B is a standard Brownian motion, g a CARMA kernel and V a Lévy subordinator independent of B. Now we replace the driving Lévy process estimated in Subsection 4.2.2 by a volatility modulated Brownian motion, so

$$dL(t) = \sigma(t)dB(t).$$

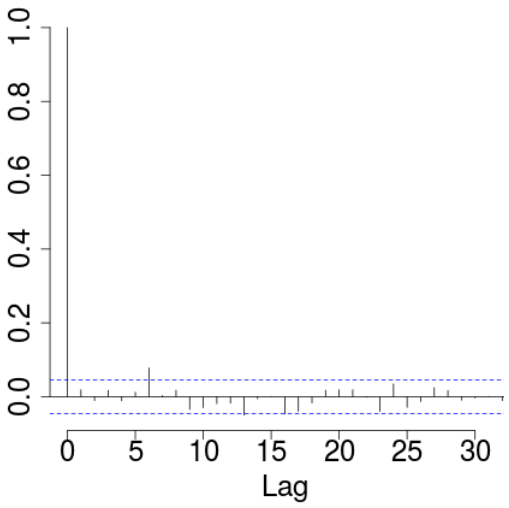


Figure 4.11: The autocorrelation function of the increments of $L(t)$.

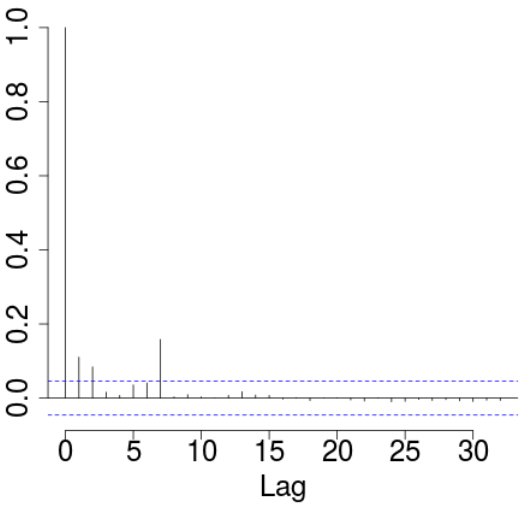


Figure 4.12: The autocorrelation function of the squared increments of $L(t)$.

Since in Chapter 7 we deseasonalised all relevant data sets, we do not explicitly introduce any seasonal components of the stochastic volatility. While one should be aware that volatility might change with seasons, we believe that this effect should not influence the results in a significant way.
For all $n = 1, \ldots, N$, where N denotes the number of observations, the estimated realised variance converges uniformly on compact sets in probability to the integrated variance $C(t)$, where

$$C(t) := \int_0^t \sigma(s)^2 ds$$

and

$$\Delta_n^h C := C_{nh} - C_{(n-1)h} = \int_{(n-1)h}^{nh} \sigma(s)^2 ds.$$

We compute:

$$\kappa_1 := \mathbb{E}_{\mathbb{Q}}\left[\sigma(t)^2\right] = \mathbb{E}_{\mathbb{Q}}\left[V(1)\right]\int_0^\infty e^{-\delta x}dx = \frac{\mathbb{E}_{\mathbb{Q}}\left[V(1)\right]}{\delta};$$

$$\kappa_2 := \mathrm{Var}_{\mathbb{Q}}\left[\sigma(t)^2\right] = \mathrm{Var}_{\mathbb{Q}}\left[V(1)\right]\int_0^\infty e^{-2\delta x}dx = \frac{\mathrm{Var}_{\mathbb{Q}}\left[V(1)\right]}{2\delta};$$

$$\mathrm{Cov}_{\mathbb{Q}}\left(\sigma(t+s)^2, \sigma(t)^2\right) = \mathrm{Var}_{\mathbb{Q}}\left[V(1)\right]\int_0^\infty e^{-\delta(x+s)}e^{-\delta x}dx$$

$$= \frac{\mathrm{Var}_{\mathbb{Q}}\left[V(1)\right]}{2\delta}e^{-\delta s} = \kappa_2 e^{-\delta s}.$$

It follows that for $k \in 0, \ldots, N-n$

$$\mathbb{E}_{\mathbb{Q}}\left[\Delta_n^h C\right] = \kappa_1;$$

$$\mathrm{Cov}_{\mathbb{Q}}\left(\Delta_{n+k}^h C, \Delta_n^h C\right) = \mathbb{E}_{\mathbb{Q}}\left[\int_{(n+k)h}^{(n+k-1)h}\sigma(s)^2 ds \int_{(n-1)h}^{nh}\sigma(u)^2 du\right] - \kappa_1^2$$

$$= \int_{(n+k)h}^{(n+k-1)h}\int_{(n-1)h}^{nh}\mathbb{E}_{\mathbb{Q}}\left[\sigma(s)^2\sigma(u)^2\right]dsdu - \kappa_1^2$$

$$= \kappa_2\int_{(n+k)h}^{(n+k-1)h}\int_{(n-1)h}^{nh}e^{-\delta|s-u|}dsdu + \kappa_1^2 - \kappa_1^2$$

$$= \kappa_2\int_{(n+k)h}^{(n+k-1)h}e^{-\delta u}du\int_{(n-1)h}^{nh}e^{\delta s}ds = \frac{\kappa_2}{\delta^2}\left(e^{\frac{\delta}{2}} - e^{-\frac{\delta}{2}}\right)^2 e^{-\delta hk};$$

$$\mathrm{Var}_{\mathbb{Q}}\left[\Delta_n^h C\right] = \mathrm{Cov}_{\mathbb{Q}}\left(\Delta_n^h C, \Delta_n^h C\right) = \frac{\kappa_2}{\delta^2}\left(e^{\frac{\delta}{2}} - e^{-\frac{\delta}{2}}\right)^2.$$

Thus for $k = 0, \ldots, N-n$ the theoretical autocorrelation function of Lévy increments converges to

$$\mathrm{ACF}(h) = e^{-\delta k}.$$

In order to estimate the memory parameter δ, we need to match this theoretical function with its empirical counterpart, both computed for

$$\sum_{n=1}^{N}\left(\Delta_n^h C\right)^2.$$

We use the function `nls` from the `R` package `stats` ([R Core Team, 2018]), with the first six lags and the starting value equal zero. It gives us an estimate of

$\hat{\delta} = 2.24$, at significance level 0.01. Figure 4.13 presents the true and estimated autocorrelation functions. We conclude that the stochastic volatility model

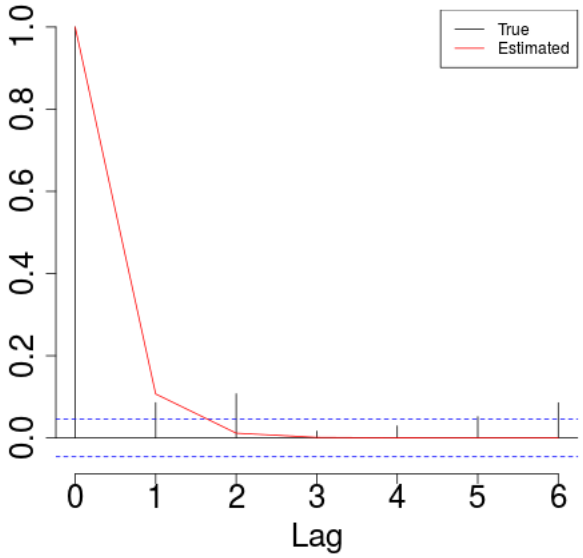


Figure 4.13: True and estimated autocorrelation functions of the squared increments of $L(t)$.

of $Y(t)$ is a valid alternative to CARMA(2,1), as indicated by correlated squared increments of the driving process $L(t)$ as well as by the presence of a statistically significant Ornstein-Uhlenbeck memory parameter. However, since the evidence of stochastic volatility in our data set is relatively weak, we expect this model to be the preferred choice in data sets with more pronounced volatility. In the remaining sections we will return to the model without stochastic volatility and propose a different modification so that we can identify the impact of different modifications on the model quality.

4.3 The influence of wind energy production on electricity prices

So far the model has not included any information about wind energy production. Therefore in this section we modify the model proposed in Sub-

section 4.1.3 and consider the relationship between the daily wind energy generation levels and spot prices in Germany and Austria between 01.01.2011 and 31.12.2015, as described in Section 3.1. We recall that because long-term weather forecasts tend to be unreliable, we assume that wind energy production influences only the short-term factor Y. Therefore Λ and Z introduced in Subsection 4.1.3 remain unchanged.

As described in Subsection 3.3.1, we explore three wind-related variables: day-ahead forecasts of wind energy generation WD, wind penetration index WPI and residual demand RD. We recall that the two latter variables are the ratio of WD to LD and the difference between LD and WD, respectively, where LD denotes the day-ahead forecast of the total load.

We start by regressing the short-term factor Y on combinations of these deseasonalised and detrended variables ($\overline{WD}$, $\overline{LD}$ and $\overline{WPI}$) and their squares. [Elberg and Hagspiel, 2015] suggest a similar approach, with some significant differences. Firstly, they only consider the residual demand, while we compare it to wind energy generation and wind penetration index. Secondly, in order to capture the non-linear relationship between spot prices and the residual demand, they suggest using splines. However, as shown in Figure 3.18, the relationships between deseasonalised wind variables and spot prices do not seem to require a polynomial level higher than quadratic. Moreover, linear regression results in a more parsimonious and intuitive model than splines, which is important for practitioners. Thirdly, instead of varying model coefficients with time like [Elberg and Hagspiel, 2015], we deseasonalise the data; again, the resulting model is easier to use in practice.

First we consider models with one wind-related variable each (models 4, 5 and 6), to check which one helps explain electricity spot prices. We then extend them by adding square terms (models 7, 8 and 9) to allow for non-linear relationships. We also look into all wind variables together (model 3) as well as all wind variables in both linear and squared forms (model 2). These two models let us compare the significance of variables of interest directly. Because the relationships between explanatory variables are not linear, we do not expect any collinearity issues. However, we include these two models to check if the impact of some variables on the prices changes due to the presence of the others. In practice we suggest using one of the remaining

models to avoid overfitting. For that reason and to keep the interpretability we do not include any models with interaction terms between explanatory variables. Finally, we consider a model without any information about the wind energy production for comparison.
In Table 4.5 and Table 4.6 we report the regression coefficients for all nine proposed models, estimated with robust linear regression and ordinary least squares, respectively. In the first case we use the `R` function rlm from the package `MASS` ([Venables and Ripley, 2002]), while in the latter `lm` from `stats` ([R Core Team, 2018]). In the remainder of the chapter we keep the coefficients estimated with robust regression, as this approach should reduce the bias towards outliers. The results of both estimation methods differ slightly, but almost only quantitatively: only $\overline{WD}$ in the model 3 is significant at level 0.05 when we use the robust regression, while insignificant with the standard linear regression.
The intercept becomes significant if and only if we include squared variables. We are particularly interested in model 2, where we include all three variables ($\overline{WD}$, $\overline{WPI}$ and $\overline{RD}$) as well as their squares. The results indicate that all three wind-related variables help explain the short-term factor $Y(t)$: $\overline{WPI}$ and $\overline{RD}$ in their basic form, while $\overline{WD}$ squared. Because $\overline{WD}^2$ and $\overline{RD}^2$ take large values, the coefficients in front of them are very small.
After regressing the short-term factor $Y(t)$ on combinations of these variables and their squares, we fit CARMA(2,1) models to the residuals.

4.3.1 Model comparisons

In this section we try to answer two questions. First, if considering wind energy data improves the model for electricity spot prices. Second, which combination of wind energy variables gives the best results.
Since statistical significance gives us valuable information about the usefulness of different wind-related variables in explaining the electricity prices, we do not remove coefficients insignificant at level 0.05. This choice allows us to interpret the compared models in a meaningful way.
We compare the models in sample, i.e. on the data set the model was fitted

	intercept	$\overline{WD}$	$\overline{WPI}$	$\overline{RD}$	$\overline{WD}^2$	$\overline{WPI}^2$	$\overline{RD}^2$
1	x	x	x	x	x	x	x
2	**-0.849**	0.0109	**-54.3**	**0.0267**	**5.28e-05**	-17.3	5.29e-06
3	-0.193	**0.0169**	**-53.9**	**0.0266**	x	x	x
4	-0.0913	**-0.0441**	x	x	x	x	x
5	-0.114	x	**-73.4**	x	x	x	x
6	-0.191	x	x	**0.038**	x	x	x
7	**-0.806**	**-0.0514**	x	x	**6.11e-05**	x	x
8	**-0.777**	x	**-84.1**	x	x	**143**	x
9	**-0.652**	x	x	**0.0409**	x	x	**2.31e-05**

Table 4.5: Coefficients of linear parts of all models of Y_t, of the form $a_0 + a_1 \cdot \overline{WD} + a_2 \cdot \overline{WPI} + a_3 \cdot \overline{RD} + a_4 \cdot \overline{WD}^2 + a_5 \cdot \overline{WPI}^2 + a_6 \cdot \overline{RD}^2$, estimated using the robust linear regression (function `rlm` in the `R` package `MASS`). With 'x' we denote variables absent in a given model. Coefficients in bold are significant at level 0.05.

to (2011–2015). As the difference between models lies only in the short-term factor $Y(t)$, we focus on this part of the model. Using the parameters estimated in Section 4.3, for each model we simulate a path of $Y(t)$ of the same length as the original data: $1,824$ days (for simulation details we refer the reader to Subsection 4.2.2). In Figure 4.14 we present the densities of the simulated and true $Y(t)$. All models replicate the density relatively well. Additionally, for all models we measure the distances between the true and simulated densities with four statistical distances provided in the `R` package `philentropy` ([Drost, 2018]):

1. Euclidean : $d = \sqrt{(\sum |P_i - Q_i|^2)}$;
2. Manhattan : $d = \sum |P_i - Q_i|$;
3. Minkowski : $d = (\sum |P_i - Q_i|^p)^{\frac{1}{p}}$;
4. Chebyshev : $d = \max |P_i - Q_i|$.

Table 4.7 presents the models whose densities are the closest and the furthest from the true $Y(t)$. The model 1 (with no wind information) maximises three out of four distances. This confirms our intuition, since the model without any wind information should explain smaller parts of prices than

	intercept	$\overline{WD}$	$\overline{WPI}$	$\overline{RD}$	$\overline{WD}^2$	$\overline{WPI}^2$	$\overline{RD}^2$
1	x	x	x	x	x	x	x
2	**-0.854**	0.00704	**-45.5**	**0.0303**	**4.19e-05**	17.9	5.93e-06
3	-0.115	0.0106	**-38.6**	**0.0302**	x	x	x
4	-0.116	**-0.0451**	x	x	x	x	x
5	-0.116	x	**-74.7**	x	x	x	x
6	-0.116	x	x	**0.0395**	x	x	x
7	**-0.912**	**-0.0542**	x	x	**6.38e-05**	x	x
8	**-0.839**	x	**-87.7**	x	x	**145**	x
9	**-0.577**	x	x	**0.0429**	x	x	**2.17e-05**

Table 4.6: Coefficients of linear parts of all models of Y_t, of the form $a_0 + a_1 \cdot \overline{WD} + a_2 \cdot \overline{WPI} + a_3 \cdot \overline{RD} + a_4 \cdot \overline{WD}^2 + a_5 \cdot \overline{WPI}^2 + a_6 \cdot \overline{RD}^2$, estimated using the ordinary linear regression (function `lm` in the `R` package `stats`). With 'x' we denote variables absent in a given model. Coefficients in bold are significant at level 0.05.

other models. On the other hand, model 9 (with $\overline{RD}$ linear and squared) minimises three out of four distances, which would confirm the insights of [Elberg and Hagspiel, 2015].

measure	1	2	3	4
min distance	9	9	9	7
max distance	1	1	1	8

Table 4.7: Best and worst models according to different statistical divergencies (in sample).

Similarly to [Cartea et al., 2009], we also compare the true and simulated summary statistics. Precisely, we simulate $1,000$ paths of $Y(t)$ of the same length as the original data: $N = 1,824$. We compute the mean (m), variance (v), skewness (s) and kurtosis (k) of each path, and average them over all paths to obtain Monte Carlo estimates of summary statistics. In Table 4.8 we present squared differences between true and Monte Carlo statistics, while in Table 4.9 the squared differences normalised by the value of the appropriate true statistics. Additionally, for all models we compute (normalised) Euclidean distances:

$$\sqrt{(\mathrm{m}^{\mathrm{true}} - \mathrm{m}^{\mathrm{MC}})^2 + (\mathrm{v}^{\mathrm{true}} - \mathrm{v}^{\mathrm{MC}})^2 + (\mathrm{s}^{\mathrm{true}} - \mathrm{s}^{\mathrm{MC}})^2 + \left(\mathrm{k}^{\mathrm{true}} - \mathrm{k}^{\mathrm{MC}}\right)^2}$$

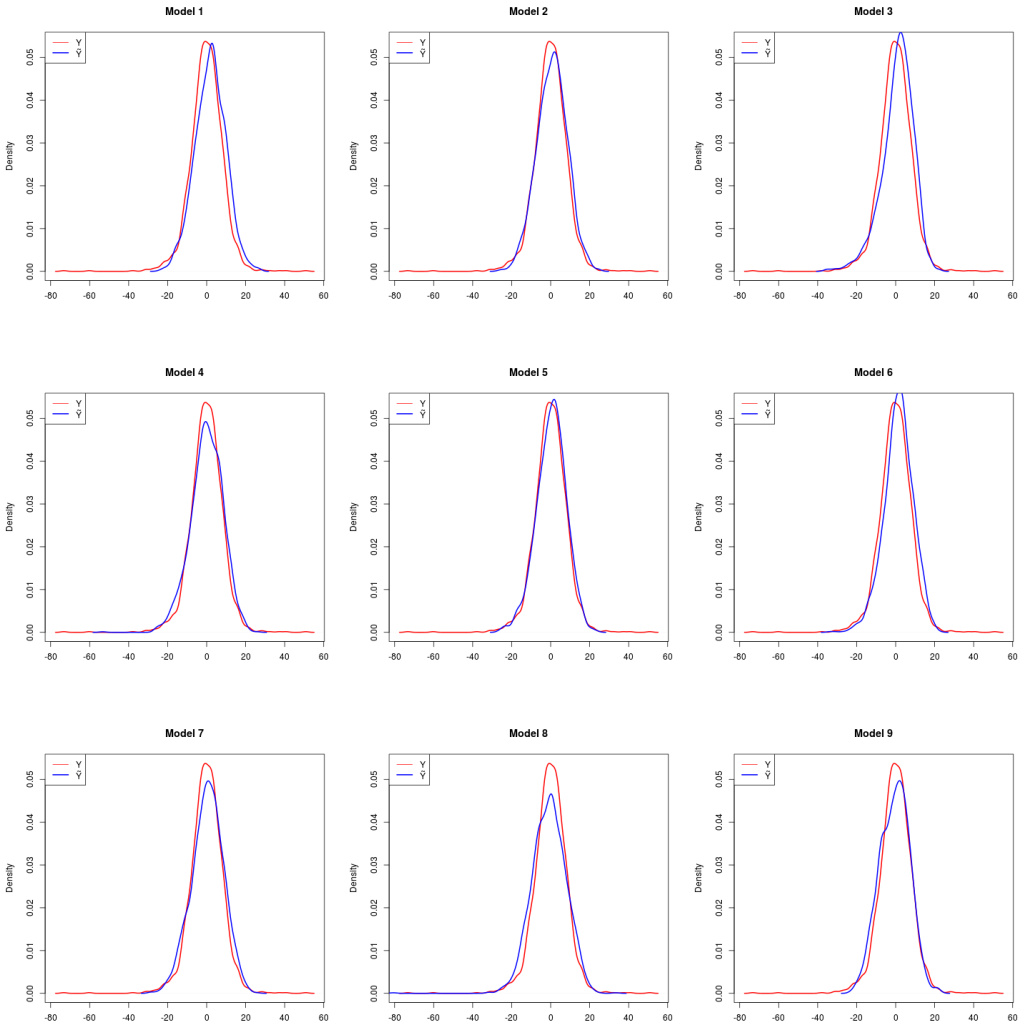


Figure 4.14: Densities of the true and simulated short-term factors $Y(t)$ for nine different model variations: in sample.

and

$$\sqrt{\left(1-\frac{\mathrm{m}^{\mathrm{MC}}}{\mathrm{m}^{\mathrm{true}}}\right)^2+\left(1-\frac{\mathrm{v}^{\mathrm{MC}}}{\mathrm{v}^{\mathrm{true}}}\right)^2+\left(1-\frac{\mathrm{s}^{\mathrm{MC}}}{\mathrm{s}^{\mathrm{true}}}\right)^2+\left(1-\frac{\mathrm{k}^{\mathrm{MC}}}{\mathrm{k}^{\mathrm{true}}}\right)^2},$$

i.e. square roots of all columns of Table 4.8 and Table 4.9. From Table 4.10 and Table 4.11 we learn that models 8 and 1 perform best according to these distances, while models 4 and 5 worst.

Model	Mean (m)	Variance (v)	Skewness (s)	Kurtosis (k)
1	0.00	14.39	0.09	17.85
2	0.09	32.45	0.13	39.41
3	0.02	25.62	[**0.02**]	35.04
4	[**0.00**]	**533.92**	**0.32**	56.07
5	**0.47**	6.81	0.05	35.99
6	0.25	24.76	0.04	35.69
7	0.00	279.02	0.29	**77.63**
8	0.01	[**0.09**]	0.06	[**14.95**]
9	0.36	43.09	0.17	40.06

Table 4.8: Squared differences between summary statistics of $Y(t)$: true and averaged over $1,000$ Monte Carlo simulations. Square brackets denote the minimum value for a given statistics, while underlines the maximum value.

Model	Mean (m)	Variance (v)	Skewness (s)	Kurtosis (k)
1	0.12	0.00	0.30	0.20
2	6.99	0.01	0.43	0.45
3	1.17	0.01	[**0.06**]	0.40
4	[**0.00**]	**0.11**	**1.05**	0.63
5	**34.71**	0.00	0.15	0.41
6	18.40	0.01	0.12	0.40
7	0.02	0.06	0.96	**0.88**
8	0.58	[**0.00**]	0.20	[**0.17**]
9	26.94	0.01	0.57	0.45

Table 4.9: Squared normalised differences between summary statistics of $Y(t)$: true and averaged over $1,000$ Monte Carlo simulations. Square brackets denote the minimum value for a given statistics, while underlines the maximum value.

Additionally, we look at the squared differences between the first ten moments of $Y(t)$: true $(\mu_1^{\mathrm{true}},\ldots,\mu_{10}^{\mathrm{true}})$ and averaged over 1,000 Monte Carlo simulations

Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Distance	5.69	8.49	7.79	**24.30**	6.58	7.79	18.89	**[3.89]**	9.15

Table 4.10: Euclidean distances between summary statistics of $Y(t)$: true and averaged over $1,000$ Monte Carlo simulations. Square brackets denote the minimum value, while underlines the maximum value.

Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Distance	**[0.79]**	2.80	1.28	1.34	**5.94**	4.35	1.38	0.97	5.29

Table 4.11: Normalised Euclidean distances between summary statistics of $Y(t)$: true and averaged over $1,000$ Monte Carlo simulations. Square brackets denote the minimum value, while the underlines the maximum value.

$(\mu_1^{\text{MC}}, \ldots, \mu_{10}^{\text{MC}})$. We also compute the same metric, but normalised by true moments. Thus for $i = 1, \ldots, 10$ we compute

$$\left(\mu_i^{\text{true}} - \mu_i^{\text{MC}}\right)^2$$

and

$$\left(1 - \frac{\mu_i^{\text{MC}}}{\mu_i^{\text{true}}}\right)^2,$$

respectively. For each distance and each number of moments we define the best model as the model minimising the distance and the worst as the one maximising the distance. We present the best and worst models in Table 4.12 (differences between moments) and Table 4.13 (sums of differences between moments). Clearly model 4 ($\overline{WD}$) has the worst performance. The best-performing models are 3 (all variables without squares) and 5 ($\overline{WPI}$).

	Best models		Worst models	
Moment	Non-normalised	Normalised	Non-normalised	Normalised
1	4	4	5	5
2	8	8	4	4
3	3	3	4	4
4	5	5	4	4
5	3	3	4	4
6	5	5	4	4
7	3	3	4	4
8	5	5	4	4
9	5	5	4	4
10	5	5	4	4

Table 4.12: Best and worst models: models minimising and maximising (normalised) squared distances between the moments of empirical and simulated data (1,000 simulations). Models are described in Table 4.5.

	Best models		Worst models	
Highest moment	Non-normalised	Normalised	Non-normalised	Normalised
1	4	4	5	5
2	8	4	5	5
3	8	8	4	5
4	5	8	4	4
5	3	3	4	4
6	5	3	4	4
7	3	3	4	4
8	5	3	4	4
9	3	3	4	4
10	5	3	4	4

Table 4.13: Best and worst models: models minimising and maximising the sum of (normalised) squared distances between the moments of empirical and simulated data (1,000 simulations), for different numbers of moments. Models are described in Table 4.5.

4.4 Modelling wind-related variables using gamma kernels

In Subsection 4.3.1 we showed that wind-related variables ($\overline{WD}$, $\overline{RD}$ and $\overline{WPI}$) help model electricity spot prices. In this section we propose models for these three variables which can be used as exogenous variables in the arithmetic model (Section 4.3) instead of empirical data.

4.4.1 Lévy semistationary processes with gamma kernels

Inspired by [Barndorff-Nielsen et al., 2013], we let

$$X(t) = \mu + c\int_{-\infty}^{t} g(t-s)\omega(s)dB(s) + \gamma\int_{-\infty}^{t} q(t-s)\omega^2(s)ds, \qquad (4.24)$$

where μ, c and γ are real constants, ω is a stationary volatility process and B denotes a standard Brownian motion independent of ω. We denote the gamma density with parameters $\bar{\nu} > 0$ and $\bar{\lambda} > 0$ by

$$\bar{g}\left(t;\bar{\nu},\bar{\lambda}\right) = \frac{\bar{\lambda}^{\bar{\nu}}}{\Gamma(\bar{\nu})}t^{\bar{\nu}-1}e^{-\bar{\lambda}t}.$$

To ensure the existence of Eq. (4.24), we assume that $\frac{1}{2} < \bar{\nu} < 1$. We define

$$g\left(t\right) = \frac{\bar{\lambda}^{\bar{\nu}-\frac{1}{2}}}{\sqrt{\Gamma(2\bar{\nu}-1)}}t^{\bar{\nu}-1}\exp\left(-\frac{\bar{\lambda}}{2}t\right)$$

and

$$q(t) = g^2(t) = \bar{g}\left(t;2\bar{\nu}-1,\bar{\lambda}\right).$$

We also let

$$\omega^2(t) = \int_{-\infty}^{t} p(t-s)dU(s)$$

for

$$p(t) = \frac{1}{\bar{\lambda}}\bar{g}\left(t;2-2\bar{\nu},\bar{\lambda}\right)$$

and a subordinator $U(t)$, which we will specify later. We observe that

$$\begin{aligned} X(t)|\omega \sim & N\left(\mu+\gamma\int_{-\infty}^{t} q(t-s)\omega^2(s)ds, c^2\int_{-\infty}^{t} g^2(t-s)\omega^2(s)ds\right) \\ =: & N\left(\mu+\gamma\sigma^2(t), c^2\sigma^2(t)\right) \end{aligned}$$

with

$$\sigma^2(t) = \int_{-\infty}^{t} e^{-\bar{\lambda}(t-s)}dU(s).$$

We can compute

$$\mathbb{E}\left[\sigma^2(0)\right] = \frac{\mathbb{E}\left[U(1)\right]}{\bar{\lambda}}$$

and

$$\mathrm{Var}\left(\sigma^2(0)\right) = \frac{\mathrm{Var}\left(U(1)\right)}{2\bar{\lambda}}.$$

[Barndorff-Nielsen et al., 2013] notice that if U denotes a subordinator such that $\sigma^2 \sim GIG(\lambda,\chi,\psi)$ then $X \sim GH\left(\lambda,\chi,\psi,\mu,c^2,\gamma\right)$, so the marginal distribution of X uniquely determines the distribution of σ^2.

Based on the explicit formula for the autocovariance of Lévy semistationary processes provided by [Barndorff-Nielsen et al., 2018, p. 24], for $s>0$ we compute the autocovariance of Eq. (4.24):

$$\begin{aligned} \mathrm{Cov}\left(X(t+s),X(t)\right) = & c^2\mathbb{E}\left[\omega^2(0)\right]\int_0^\infty g(x+s)g(x)dx \\ & +\gamma^2\int_0^\infty\int_0^\infty q(x+s)q(y)\mathrm{Cov}\left(\omega^2(|x-y|),\omega^2(0)\right)dxdy. \end{aligned}$$

Let us study each term separately. By [Barndorff-Nielsen et al., 2013],

$$\begin{aligned} \mathbb{E}\left[\omega^2(0)\right] &= \mathbb{E}\left[U(1)\right]\int_0^\infty p(x)dx \\ &= \mathbb{E}\left[U(1)\right]\int_0^\infty \frac{1}{\bar{\lambda}}\frac{\bar{\lambda}^{2-2\bar{\nu}}}{\Gamma\left(2-2\bar{\nu}\right)}x^{1-2\bar{\nu}}e^{-\bar{\lambda}x}dx = \frac{\mathbb{E}\left[U(1)\right]}{\bar{\lambda}} = \mathbb{E}\left[\sigma^2(0)\right] \end{aligned}$$

and

$$\begin{aligned}&\mathrm{Cov}\left(\omega^2(t+s),\omega^2(t)\right)=\mathrm{Var}\left(U(1)\right)\int_0^\infty p(x+s)p(x)dx\\&=2\bar{\lambda}\mathrm{Var}\left(\sigma^2(0)\right)\int_0^\infty p(x+s)p(x)dx.\end{aligned}$$

Furthermore,

$$\begin{aligned}&\int_0^\infty g(x+s)g(x)dx=\int_0^\infty\frac{\bar{\lambda}^{2\nu-1}}{\Gamma\left(2\bar{\nu}-1\right)}\left((x+s)x\right)^{\bar{\nu}-1}e^{-\frac{\bar{\lambda}}{2}(2x+s)}dx\\&=\frac{1}{\Gamma(\bar{\nu}-\frac{1}{2})}2^{\frac{3}{2}-\bar{\nu}}\bar{K}_{\bar{\nu}-\frac{1}{2}}\left(\frac{\bar{\lambda}}{2}s\right),\end{aligned}$$

where

$$\bar{K}_\alpha\left(x\right)=x^\alpha K_\alpha\left(x\right)$$

and $K_\nu\left(x\right)$ denotes the modified Bessel function of the third kind. Finally,

$$\begin{aligned}&\int_0^\infty p(x+s)p(x)dx=\int_0^\infty\left(\frac{1}{\bar{\lambda}}\frac{\bar{\lambda}^{2-2\bar{\nu}}}{\Gamma\left(2-2\bar{\nu}\right)}\right)^2\left((x+s)x\right)^{1-2\bar{\nu}}e^{-\bar{\lambda}(2x+s)}dx\\&=\frac{1}{\bar{\lambda}^3}\left(\frac{1}{\Gamma\left(2-2\bar{\nu}\right)}\right)^2\frac{\Gamma\left(3-2\bar{\nu}\right)}{\Gamma\left(\frac{1}{2}\right)}2^{2\bar{\nu}-\frac{5}{2}}\bar{K}_{\nu-\frac{1}{2}}\left(\bar{\lambda}s\right).\end{aligned}$$

Therefore

$$\mathrm{Cov}\left(X(t+s),X(t)\right)=P+Q,$$

with

$$P=c^2\mathbb{E}\left[\sigma^2(0)\right]\frac{2^{\frac{3}{2}-\bar{\nu}}}{\Gamma(\bar{\nu}-\frac{1}{2})}\bar{K}_{\bar{\nu}-\frac{1}{2}}\left(\frac{\bar{\lambda}}{2}s\right)$$

and

$$\begin{aligned}&Q=\gamma^2 2\bar{\lambda}\mathrm{Var}\left(\sigma^2(0)\right)\frac{1}{\bar{\lambda}^3}\left(\frac{1}{\Gamma\left(2-2\bar{\nu}\right)}\right)^2\frac{\Gamma\left(3-2\bar{\nu}\right)}{\Gamma\left(\frac{1}{2}\right)}2^{2\bar{\nu}-\frac{5}{2}}\frac{\bar{\lambda}^{4\bar{\nu}-2}}{\Gamma\left(2\bar{\nu}-1\right)^2}\\&\int_0^\infty\int_0^\infty(x+s)^{2\bar{\nu}-2}e^{-\bar{\lambda}(x+s)}y^{2\bar{\nu}-2}e^{-\bar{\lambda}y}\bar{K}_{\bar{\nu}-\frac{1}{2}}\left(\bar{\lambda}|x-y|\right)dxdy\\&=c_Q\int_0^\infty\int_0^\infty(x+s)^{2\bar{\nu}-2}e^{-\bar{\lambda}(x+s)}y^{2\bar{\nu}-2}e^{-\bar{\lambda}y}\bar{K}_{\bar{\nu}-\frac{1}{2}}\left(\bar{\lambda}|x-y|\right)dxdy,\end{aligned}$$

where

$$c_Q = \gamma^2 \bar{\lambda}^{4\bar{\nu}-4} \mathrm{Var}\left(\sigma^2(0)\right) 2^{2\bar{\nu}-\frac{3}{2}} \frac{\Gamma\left(3-2\bar{\nu}\right)}{\Gamma\left(\frac{1}{2}\right)} \left(\frac{1}{\Gamma\left(2-2\bar{\nu}\right)\Gamma\left(2\bar{\nu}-1\right)}\right)^2.$$

Similarly, we compute the variance of X:

$$\mathrm{Var}\left(X(0)\right) = P_0 + Q_0,$$

with

$$P_0 = c^2 \mathbb{E}\left[\sigma^2(0)\right]$$

and

$$Q_0 = c_Q \int_0^\infty \int_0^\infty (xy)^{2\bar{\nu}-2} e^{-\bar{\lambda}(x+y)} \bar{K}_{\bar{\nu}-\frac{1}{2}}\left(\bar{\lambda}|x-y|\right) dxdy.$$

Now we are ready to estimate the parameters $\bar{\lambda}$ and $\bar{\nu}$ of the gamma kernel by matching the first lags of the empirical and theoretical autocorrelation functions, with the latter equal to

$$\mathrm{cor}\left(X(t), X(t+s)\right) = \frac{P+Q}{P_0+Q_0}.$$

4.4.2 Estimation procedure

We start the estimation procedure by using the function `stepAIC.ghyp` from the `R` package `ghyp` ([Lüthi and Breymann, 2016]) to fit 11 generalised hyperbolic distributions to X, where X denotes the wind variable of interest ($\overline{WD}$, $\overline{RD}$ or $\overline{WPI}$). As usually, we find the distribution which minimises the Akaike information criterion (AIC). Finally, we convert the parameters of the best fitting marginal generalised hyperbolic distribution (Section A.3) to the parametrisation $\left(\lambda, \chi, \psi, \mu, c = \sqrt{\Sigma}, \gamma\right)$.
After fitting the marginal distribution, we proceed to estimate kernel parameters $\bar{\nu}$ and $\bar{\lambda}$ by matching first lags of the empirical and theoretical autocorrelation functions of $X(t)$. [Barndorff-Nielsen et al., 2013] suggest using the first $\left\lfloor\sqrt{N}\right\rfloor$ lags, where N denotes the number of observations. Therefore we apply the function `optim` from the `R` package `stats` ([R Core Team, 2018])

to minimise the squared difference between the first $\lfloor\sqrt{1824}\rfloor = 42$ lags of the empirical and theoretical autocorrelation functions. We set the initial parameters to $\bar{\nu} = 0.99$ and $\bar{\lambda} = 0.01$.

Remark 4.3. The optimisation procedure is not sensitive to reasonable starting values. For example, setting the initial values to $\bar{\nu} = 0.75$ and $\bar{\lambda} = 0.20$ results in the same optimal parameters up to the third decimal place. On the other hand, the number of lags used in the estimation procedure influences the results. In case of the residual demand, the estimation procedure performed on six lags, with the original initial values, resulted in $\bar{\nu} = 0.87$ and $\bar{\lambda} = 0.26$ instead of $\bar{\nu} = 0.82$ and $\bar{\lambda} = 0.20$. This should not surprise us considering the changing behaviour of the autocorrelation function. Therefore the choice of the number of lags must correspond to the application of interest.

4.4.3 Numerical results

In this subsection we describe the results of the estimation procedure from Subsection 4.4.2 run on three wind-related variables.

4.4.3.1 Wind energy production

In Table 4.14 we can see that the asymmetric normal inverse Gaussian distribution minimises the Akaike information criterion (AIC), therefore it provides the best fit. Figure 4.15 shows that the fit is indeed very good, apart from some imperfections in the right tail. Therefore we assume that the marginal distribution of $\overline{WD}$ is asymmetric hyperbolic with

$$\left(\lambda, \chi, \psi, \mu, c = \sqrt{\Sigma}, \gamma\right) = (-0.50, 2.21, 2.21, -121.45, 8.79, 121.39)\,.$$

The value of the parameter γ is positive, which reflects the positive skewness of $\overline{WD}$.

For the kernel we obtain parameter estimates $\bar{\nu} = 0.99$ and $\bar{\lambda} = 0.57$, so similarly to [Barndorff-Nielsen et al., 2013] we are outside the semimartingale setting, but this does not cause any problems since we are not modelling

	model	symmetric	lambda	alpha.bar	mu	sigma	gamma	aic	llh
3	NIG	FALSE	-0.500	2.212	-121.451	77.228	121.395	21945.436	-10968.718
5	t	FALSE	-6.690	0.000	-212.251	58.792	212.452	21948.790	-10970.395
2	hyp	FALSE	1.000	1.682	-113.757	79.374	113.742	21950.036	-10971.018
1	ghyp	FALSE	-6.715	0.000	-213.113	58.569	213.321	21950.789	-10970.394
4	VG	FALSE	2.067	0.000	-101.971	83.414	101.966	21962.771	-10977.386
8	NIG	TRUE	-0.500	0.716	-20.367	113.947	0.000	22192.167	-11093.084
6	ghyp	TRUE	-0.524	0.715	-20.333	113.970	0.000	22194.166	-11093.083
7	hyp	TRUE	1.000	0.331	-20.978	112.161	0.000	22197.557	-11095.779
10	t	TRUE	-1.802	0.000	-18.542	118.648	0.000	22198.139	-11096.070
9	VG	TRUE	1.171	0.000	-21.571	112.311	0.000	22199.454	-11096.727
11	gauss	TRUE		Inf	-0.000	111.723	0.000	22383.348	-11189.674

Table 4.14: Generalised hyperbolic distributions fitted to $\overline{WD}$ with parametrisation $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$.

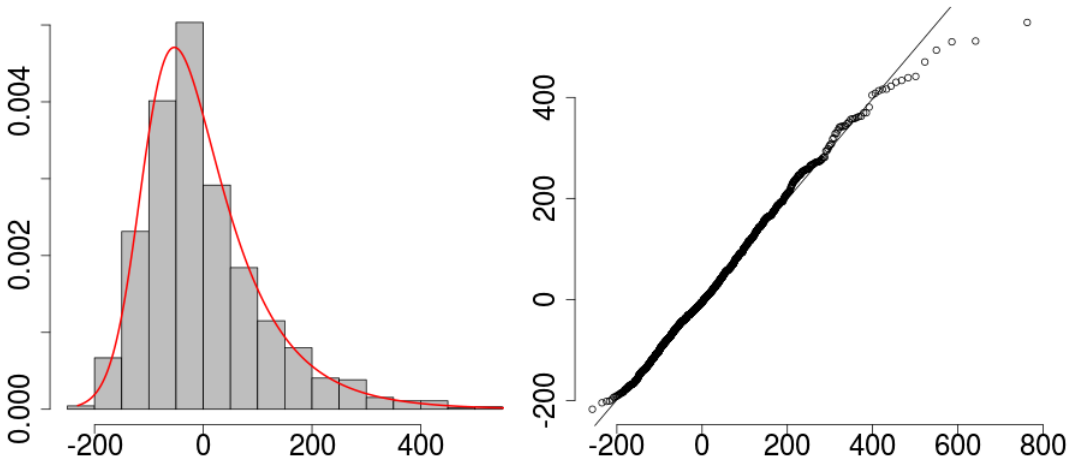


Figure 4.15: $\overline{WD}$ and the fitted generalised hyperbolic distribution: histogram and q-q plot.

a tradable asset. The fit between autocorrelation functions is good, as shown in Figure 4.16.

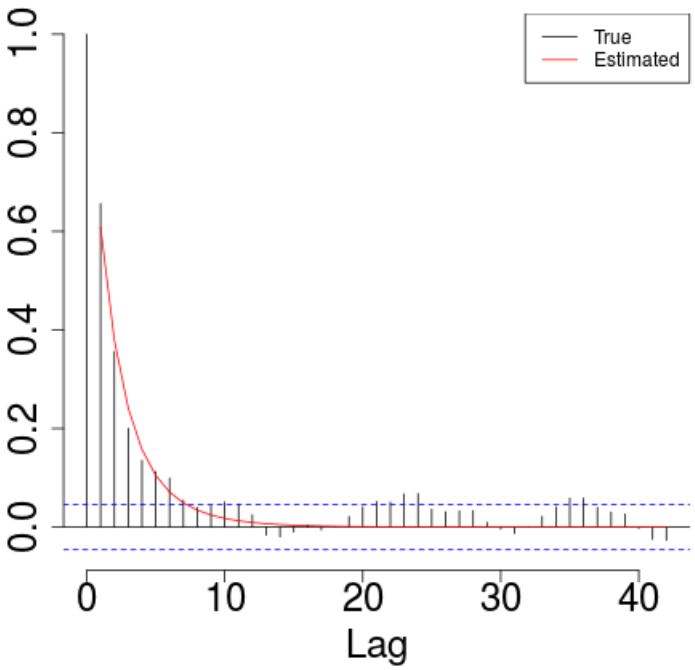


Figure 4.16: True and estimated autocorrelation functions of $\overline{WD}$ according to the model described in Section 4.4.

4.4.3.2 Residual demand

A quick look at Figure 3.18 suggests that we expect the marginal distribution of $\overline{RD}$ to be asymmetric. Table 4.15 shows that indeed the best fit is provided by the asymmetric hyperbolic distribution. Apart from some minor issues in the left tail, this distribution fits very well, as we observe in Figure 4.17. Therefore the marginal distribution of $\overline{RD}$ is asymmetric hyperbolic with

$$\left(\lambda, \chi, \psi, \mu, c = \sqrt{\Sigma}, \gamma\right) = (1.00, 0.26, 2.53, 96.44, 11.12, -96.43)\,.$$

	model	symmetric	lambda	alpha.bar	mu	sigma	gamma	aic	llh
2	hyp	FALSE	1.000	0.809	96.439	123.598	-96.435	23019.220	-11505.610
4	VG	FALSE	1.568	0.000	95.617	124.152	-95.607	23019.521	-11505.760
1	ghyp	FALSE	1.253	0.595	95.796	123.875	-95.795	23021.045	-11505.522
3	NIG	FALSE	-0.500	1.532	108.093	120.079	-108.117	23024.637	-11508.319
5	t	FALSE	-4.000	0.000	149.128	112.832	-150.102	23041.408	-11516.704
8	NIG	TRUE	-0.500	0.898	21.547	147.663	0.000	23199.988	-11596.994
6	ghyp	TRUE	-0.559	0.895	21.473	147.727	0.000	23201.982	-11596.991
7	hyp	TRUE	1.000	0.655	21.368	145.660	0.000	23205.129	-11599.565
10	t	TRUE	-1.996	0.000	19.800	152.178	0.000	23205.779	-11599.889
9	VG	TRUE	1.475	0.000	21.750	145.478	0.000	23209.951	-11601.976
11	gauss	TRUE		Inf	-0.000	145.640	0.000	23350.474	-11673.237

Table 4.15: Generalised hyperbolic distributions fitted to $\overline{RD}$ with parametrisation $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$.

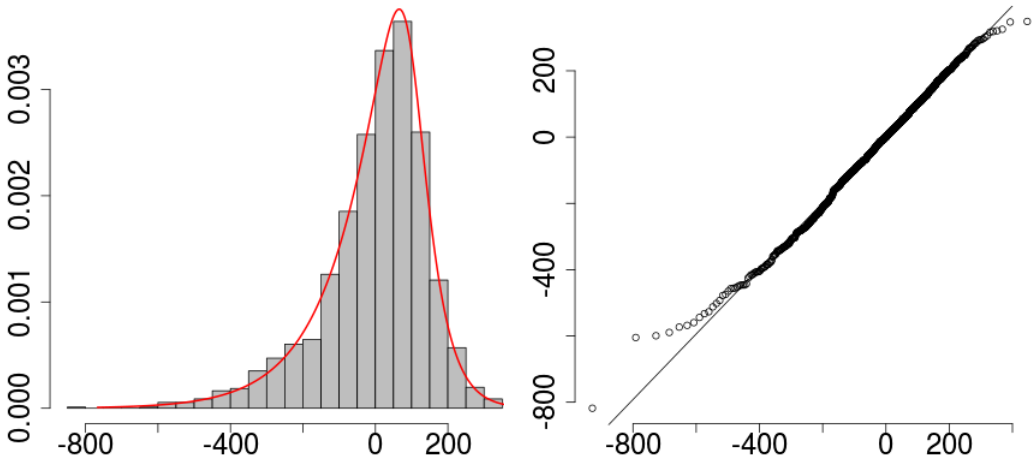


Figure 4.17: $\overline{RD}$ and the fitted generalised hyperbolic distribution: histogram and q-q plot.

The kernel parameters are estimated as $\bar{\nu} = 0.82$ and $\bar{\lambda} = 0.20$, so again we are not dealing with a semimartingale. Figure 4.18 presents the good match between the empirical and theoretical autocorrelation functions.

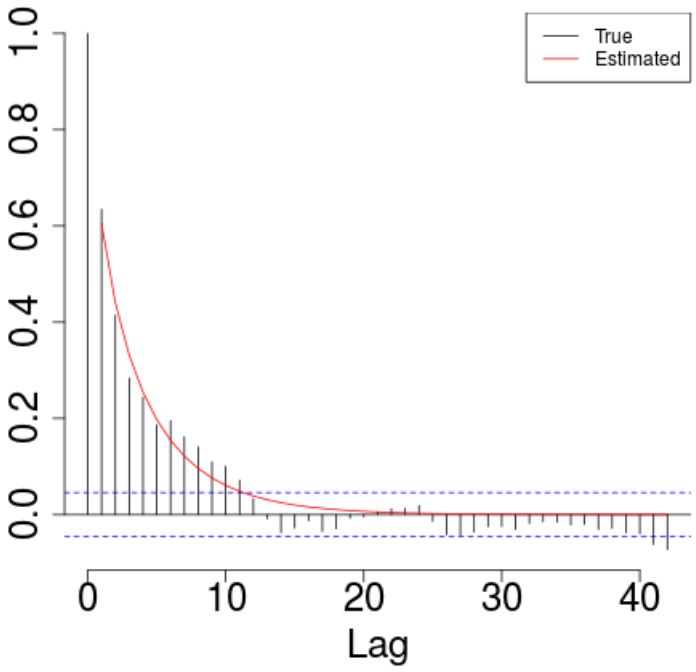


Figure 4.18: True and estimated autocorrelation functions of $\overline{RD}$ according to the model described in Section 4.4.

4.4.3.3 Wind penetration index

Similarly to $\overline{WD}$, the deseasonalised wind penetration index is best described by the asymmetric normal inverse Gaussian distribution. From Table 4.16 we learn that the skewness parameter γ is positive, but of smaller magnitude than the one of $\overline{WD}$. Again, we are satisfied with the fit presented in Figure 4.19. The marginal distribution of $\overline{WPI}$ is thus asymmetric hyperbolic with

$$\left(\lambda, \chi, \psi, \mu, c = \sqrt{\Sigma}, \gamma\right) = (-0.50, 1.71, 1.71, -0.07, 0.22, 0.07)\,.$$

	model	symmetric	lambda	alpha.bar	mu	sigma	gamma	aic	llh
3	NIG	FALSE	-0.500	1.712	-0.067	0.050	0.067	-4956.436	2482.218
2	hyp	FALSE	1.000	1.110	-0.062	0.052	0.062	-4955.579	2481.789
1	ghyp	FALSE	0.227	1.447	-0.063	0.051	0.063	-4955.566	2482.783
4	VG	FALSE	1.757	0.000	-0.059	0.053	0.059	-4945.900	2476.950
5	t	FALSE	-4.411	0.000	-0.095	0.045	0.095	-4940.267	2474.133
8	NIG	TRUE	-0.500	0.686	-0.014	0.072	0.000	-4688.184	2347.092
6	ghyp	TRUE	-0.618	0.676	-0.014	0.072	0.000	-4686.234	2347.117
10	t	TRUE	-1.750	0.000	-0.012	0.076	0.000	-4681.876	2343.938
7	hyp	TRUE	1.000	0.402	-0.014	0.071	0.000	-4678.876	2342.438
9	VG	TRUE	1.206	0.000	-0.014	0.071	0.000	-4674.755	2340.378
11	gauss	TRUE		Inf	0.000	0.071	0.000	-4493.115	2248.557

Table 4.16: Generalised hyperbolic distributions fitted to $\overline{WPI}$ with parametrisation $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$.

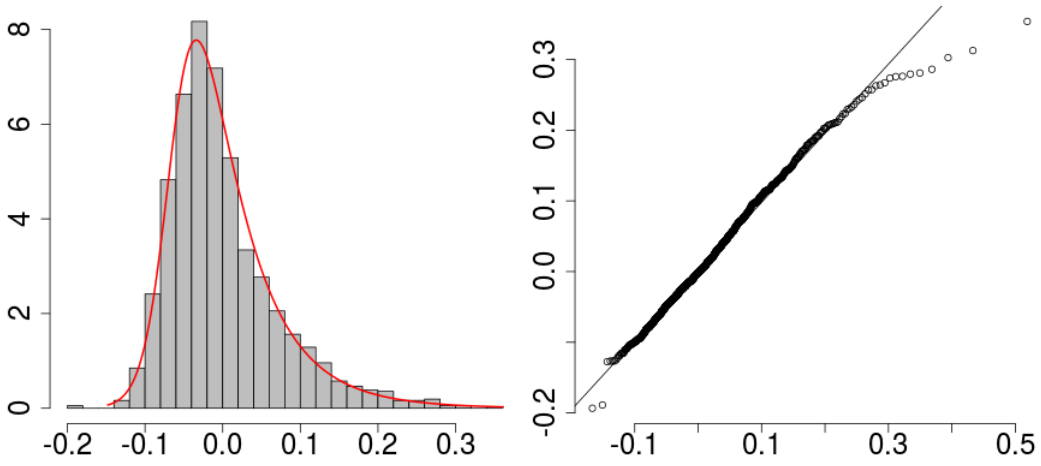


Figure 4.19: $\overline{WPI}$ and the fitted generalised hyperbolic distribution: histogram and q-q plot.

For the kernel we obtain the parameter estimates $\bar{\nu} = 0.88$ and $\bar{\lambda} = 0.39$, so once again we are outside the semimartingale setting. These parameters provide good matching between autocorrelation functions, as presented in Figure 4.20.

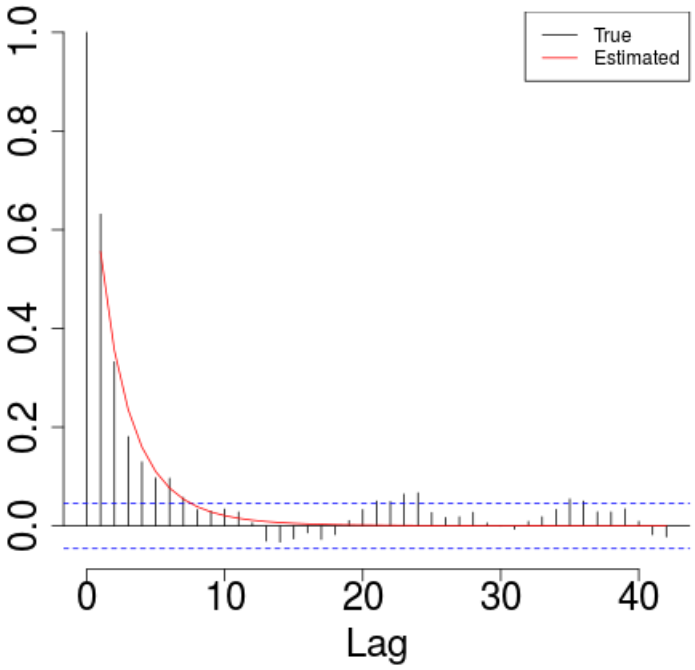


Figure 4.20: True and estimated autocorrelation functions of $\overline{WPI}$ according to the model described in Section 4.4.

4.5 Summary and outlook

In this chapter we introduced a three-factor arithmetic model of electricity spot prices, consisting of a deterministic seasonality and trend function as well as short- and long-term stochastic components, and derived a formula for futures prices. We modelled the long-term component as a Lévy process with increments belonging to the class of generalised hyperbolic distributions. We described the short-term factor with Lévy semistationary processes. We started from a CARMA(2,1) model, and generalised it by adding a short-

memory stochastic volatility. We further modified the model by including the information about the wind energy production and total load as exogenous variables. We fitted our models to Austrian and German data including spot and futures prices as well as the wind energy production and total load data. Empirical studies revealed that the basic model (without wind energy production data or stochastic volatility) already provides a good fit to the data. We compared this benchmark model with alternatives including wind energy generation. Each suggested model has different benefits and drawbacks, so the choice should depend on the application of interest. For example, the model including residual demand in the linear and squared form (model 9) minimises the distances between true and fitted densities according to three out of four tested metrics. On the other hand, the model without the information about wind energy production (model 1) maximises the same distances. If the quality of interest is the distances between empirical and fitted summary statistics, we suggest using the model with wind penetration index in the linear and squared form (model 8). If one is interested in the goodness of fit of higher moments, the model with wind penetration index (model 5) or the model with all wind variables (model 3) would both be good options.

We rounded the chapter off by modelling wind-related variables – wind energy production, wind penetration index and residual demand – with Lévy semistationary processes with gamma kernels. The empirical results showed that this type of model suits our application very well. Models of wind-related variables enable practitioners to apply models of electricity prices described in Section 4.3 without relying on additional data. In Chapter 5 we will extend this idea and propose a joint model of wind energy generation and electricity prices.

Our studies indicate that the inclusion of wind energy production data in electricity pricing models is a promising area of research. We expect this effect to grow as wind energy constitutes an increasingly bigger part of total energy production in many countries – therefore its impact on the prices will increase. Furthermore, we suggest repeating similar experiments with other renewables such as solar, possibly making use of already existing models of solar power generation ([Lingohr and Müller, 2019] or [Veraart and Zdanowicz, 2015]). Since the participation of solar power in the total energy generation

in Austria and Germany is still smaller than of the wind power, its impact might be smaller. However, the inclusion of other renewables would give us the full picture and help explain the prices even better due to their priority in the electricity market (merit-order effect).

5

A joint model of wind energy production and electricity spot prices

In Chapter 4 we found that wind energy production, possibly together with total load, helps explain the electricity spot prices. So far we included these additional information as exogenous variables. In this chapter we aim to propose a joint model of electricity spot prices and wind energy production. We recall that correlation coefficients between the deseasonalised electricity spot prices and wind related variables – deseasonalised wind energy production forecasts, deseasonalised wind penetration index forecasts and deseasonalised residual demand forecasts – equal -0.60, -0.63 and 0.72, respectively. Our joint model should account for such a strong relationship.

A joint model of wind energy production and energy prices has important applications. First, it can provide a better way of modelling both variables because, as opposed to modelling them separately, it includes the crucial information about the relationship between them. Furthermore, to trade some contracts we must be able to compute expected values not only of both variables but also the product between them. Here we explore an example of such contract introduced by [Deschatre and Veraart, 2017].

We fit a discrete vector autoregressive model to the wind energy production and electricity prices data from 2011–2015 described in Chapter 3, after necessary transformations. We then test the model on the data from 2016. Since the intended application involves trading energy contracts, in the validation we focus on one of the most important properties for practitioners: the tail behaviour. Therefore we study if our model is able to provide the information about the risk connected to wind energy generation and trading. We also prove that the joint model performs better than an alternative model which assumes the independence of variables.

5.1 The multivariate autoregressive model

We suggest modelling wind energy production and spot prices with a discrete (bivariate) vector autoregressive process. We consider two variables described in Chapter 3: electricity spot prices $S(t)$ and wind energy production day-ahead forecasts $WD(t)$, both from 2011 to 2015. We assume that for $t = 2, 3, \ldots$

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

where Λ^S is the trend and seasonality function described in Chapter 3 (please note that here Y denotes a different process than in Chapter 4). However, we modify the deseasonalising procedure of the wind energy production to ensure that simulated WD takes only non-negative values. For $t = 2, 3, \ldots$ we let

$$\sqrt{WD(t)} = \Lambda^{\sqrt{WD}}(t) + X(t),$$

where, similarly to Λ^{WD} described in Chapter 3, the seasonality function has the form

$$\Lambda^{\sqrt{WD}}(t) = c_0^{\sqrt{WD}} + c_1^{\sqrt{WD}} t + c_2^{\sqrt{WD}} \cos\left(\frac{c_3^{\sqrt{WD}} + 2\pi t}{365}\right).$$

For the details of the estimation procedure we refer the reader to Chapter 3.

Remark 5.1. We could transform wind energy production data in many alternative ways described jointly as Box-Cox transformations with a general

formula

$$f(x;\lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(x), & \lambda \neq 0. \end{cases}$$

The function `BoxCox.lambda` from the `R` package `forecast` ([Hyndman and Khandakar, 2008]) uses the maximum likelihood method to find the optimal λ. In our case it resulted in $\lambda = 0.07$, suggesting a transformation close to the natural logarithm. However, we decided to use the square root instead, for two main reasons. First, its inverse transformation does not magnify small discrepancies between the simulated and empirical data as much as exponentiating does. Second, the model with log-transformed data does not provide an intuitive interpretation. For example, when we use the logarithm, Y Granger-causes X (cf. below), even though Y does not Granger-cause $\exp(X)$. In other words, such a model would suggest spurious relationships between variables.

As an initial test of the relationship between X and Y we check the Granger causality. It is the common way of finding out if the past of one stationary variable influences another stationary variable. First we decide the number of relevant lags, in this case 2. To establish if X Granger-causes Y, for $t = 2, 3, \ldots$ we fit two regression models to Y:

$$Y(t) = a_0 + a_1 Y(t-1) + a_2 Y(t-2) + \varepsilon(t)$$

and

$$Y(t) = a_0 + a_1 Y(t-1) + a_2 Y(t-2) + b_1 X(t-1) + b_2 X(t-2) + \varepsilon(t).$$

We test the null hypothesis that $b_1 = b_2 = 0$, i.e. that past lags of X do not help model Y. We stress that Granger causality corresponds to forecasting ability, not true causality.
The Granger test performed with the function `grangertest` from the `R` package `lmtest` ([Zeileis and Hothorn, 2002]) with two lags confirms (on significance level 0.001) that X Granger-causes Y but Y does not Granger-cause Y, as expected: it is the wind energy production which influences electricity prices, not vice versa. This suggests that a vector autoregressive

model could be a good choice. For $t \in \{2, 3, \dots\}$ we choose the process VAR(2)

$$\begin{aligned} X(t) &= \beta_{X0} + \beta_{X1}Y(t-1) + \beta_{X2}Y(t-2) + \gamma_{X1}X(t-1) + \gamma_{X2}X(t-2) + \varepsilon_X(t); \\ Y(t) &= \beta_{Y0} + \beta_{Y1}Y(t-1) + \beta_{Y2}Y(t-2) + \gamma_{Y1}X(t-1) + \gamma_{Y2}X(t-2) + \varepsilon_Y(t). \end{aligned} \tag{5.1}$$

In Table 5.1 and Table 5.2 we present coefficients estimated with the `R` package `vars` ([Pfaff, 2008]). We observe that while Y has no statistically significant influence on X, X impacts Y. If we assume higher orders of the VAR model, X is influenced only by its two previous lags, while Y must include lags of both variables. Since two lags suffice to model the wind energy production, we do not increase the number of lags in the model Eq. (5.1). Figure 5.1 and Figure 5.2 suggest a very mild stochastic volatility in both

	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
γ_{X1}	**0.702**	0.029	24.071	0.000
β_{X1}	-0.020	0.015	-1.324	0.186
γ_{X2}	**-0.108**	0.029	-3.709	0.000
β_{X2}	0.020	0.015	1.307	0.191
β_{X0}	0.001	0.072	0.015	0.988

Table 5.1: Parameters of X modelled by Eq. (5.1) estimated using the `R` package `vars`. In bold: coefficients estimated on the confidence level 0.001.

	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
γ_{Y1}	-0.077	0.057	-1.358	0.175
β_{Y1}	**0.578**	0.029	19.750	0.000
γ_{Y2}	**0.214**	0.057	3.749	0.000
β_{Y2}	**0.132**	0.029	4.519	0.000
β_{Y0}	-0.006	0.141	-0.040	0.968

Table 5.2: Parameters of Y modelled by Eq. (5.1) estimated using the `R` package `vars`. In bold: coefficients estimated on the confidence level 0.001.

processes. From Figure 5.3 we learn that the residuals of X and Y are uncorrelated beyond lag 0, where the correlation equals -0.61.

We model the residuals with a distribution belonging to the bivariate generalised hyperbolic class (for more details we refer the reader to [Lüthi and

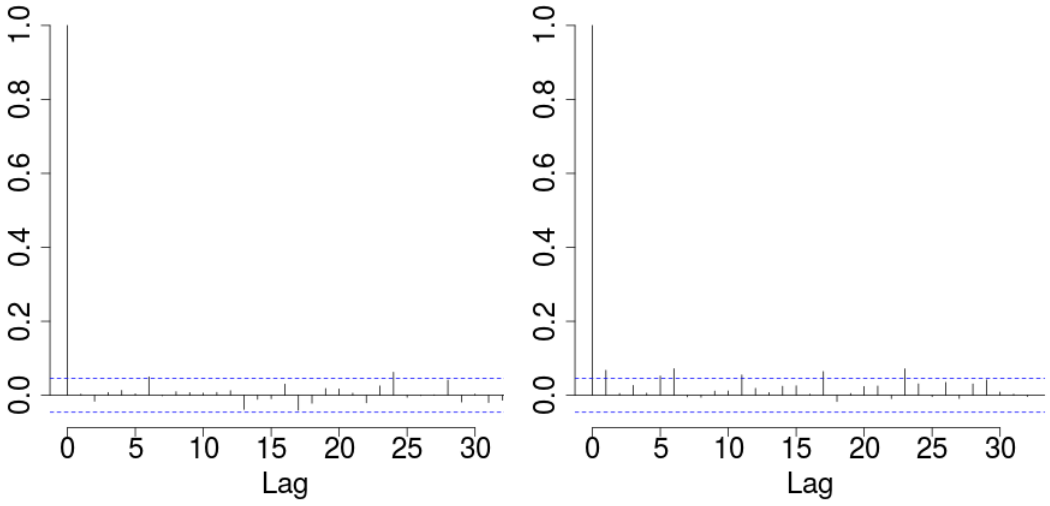


Figure 5.1: Residuals (left) and squared residuals (right) of the VAR(2) model fitted to $X(t)$.

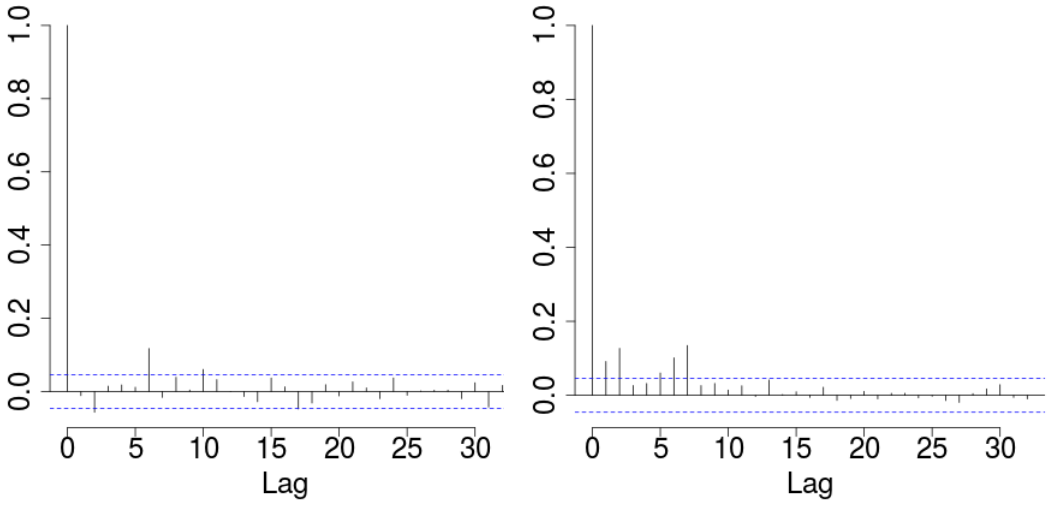


Figure 5.2: Residuals (left) and squared residuals (right) of the VAR(2) model fitted to $Y(t)$.

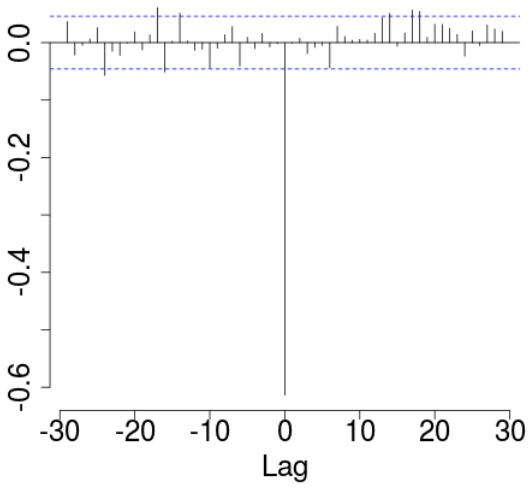


Figure 5.3: The cross-correlation between the residuals of X and Y, i.e. the autocorrelation between $\varepsilon_{X(t+k)}$ and ε_{Yt} for lags $k \in \mathbb{Z}$.

Breymann, 2016]). We use the `R` package `ghyp` to rank 11 distributions from this class according to Akaike information criterion, as shown in Table 5.3. The best fit is provided by the asymmetric Student's-t distribution with $\lambda = -2.745$, which corresponds to 5.49 degrees of freedom. This means that the process requires much heavier tails than the bivariate normal distribution. Figure 5.4 shows the qq-plots of the fitted marginals as well as the the empirical scatter plot and empirical bivariate histogram.

For comparison we fit univariate generalised hyperbolic distributions to ε_X and ε_Y, following the same procedure, where we assume uncorrelated residuals. The best fit to the first variable is provided by the asymmetric variance-gamma, while to the latter by asymmetric Student's-t distribution (with 4.39 degrees of freedom). Figure 5.5 and Figure 5.6 show that modelling both variables separately could improve the individual fits. However, this approach would neglect the clear evidence of their strong relationship presented in Figure 5.4. Instead of the bivariate generalised hyperbolic we could also jointly model the residuals with a bivariate Gaussian distribution, which is the most common assumption of vector autoregressive processes. To compare these three approaches we simulate three sets of two-dimensional points on

the plane corresponding to true residuals (of length 1826 equal to 5 years of the data). We compute the Euclidean distances between true and simulated two-dimensional residuals and present the summary statistics of the results in Table 5.4. The bivariate generalised hyperbolic model minimises the quartiles as well as the median of the distance distribution. The worst performing model is the univariate one, which indicates the assumption of correlated residuals improves the fit.

	model	symmetric	lambda	alpha.bar	aic	llh
5	t	FALSE	-2.745	0.000	19651.496	-9817.748
1	ghyp	FALSE	-2.750	0.000	19653.493	-9817.747
3	NIG	FALSE	-0.500	1.566	19676.652	-9830.326
10	t	TRUE	-2.644	0.000	19679.825	-9833.912
6	ghyp	TRUE	-2.643	0.000	19681.825	-9833.913
2	hyp	FALSE	1.500	1.280	19700.297	-9842.148
4	VG	FALSE	2.221	0.000	19710.961	-9847.481
8	NIG	TRUE	-0.500	1.490	19715.531	-9851.766
7	hyp	TRUE	1.500	1.302	19746.377	-9867.189
9	VG	TRUE	2.226	0.000	19760.620	-9874.310
11	gauss	TRUE		Inf	20160.802	-10075.401

Table 5.3: Bivariate generalised hyperbolic distributions fitted to the residuals of $X(t)$ and $Y(t)$ with parametrisation $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Bivariate ghyp	0.31	3.68	6.06	7.45	9.86	64.97
Univariate ghyp	0.39	4.35	6.81	7.94	10.22	65.74
Bivariate Gaussian	0.12	3.99	6.40	7.74	10.11	62.68

Table 5.4: Summary statistics of Euclidean distances between residuals $(\varepsilon_X, \varepsilon_Y)$ and three fitted processes: bivariate generalised hyperbolic, univariate generalised hyperbolic and bivariate Gaussian.

Finally, in Figure 5.7 we present the reaction of variables to a shock from the other variable called the impulse response function, based on the Wold (moving average) representation of the VAR process ([Pfaff, 2008]). The impulse responses were computed by the function `irf` from the `R` package `vars` ([Pfaff, 2008]). At time 0 we apply the shock to one of the variables at time and study the reaction of the other series from time 0 to 20. The red dotted

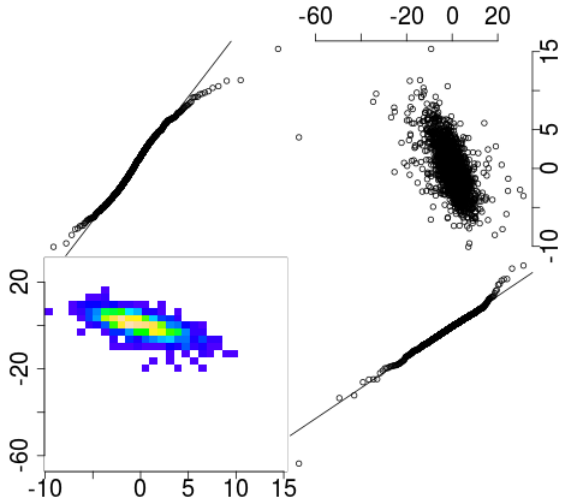


Figure 5.4: The qq-plots of the marginals fitted to the residuals of $X(t)$ (upper left) and $Y(t)$ (bottom right) as well as the bivariate histogram and the scatter plot of true residuals.

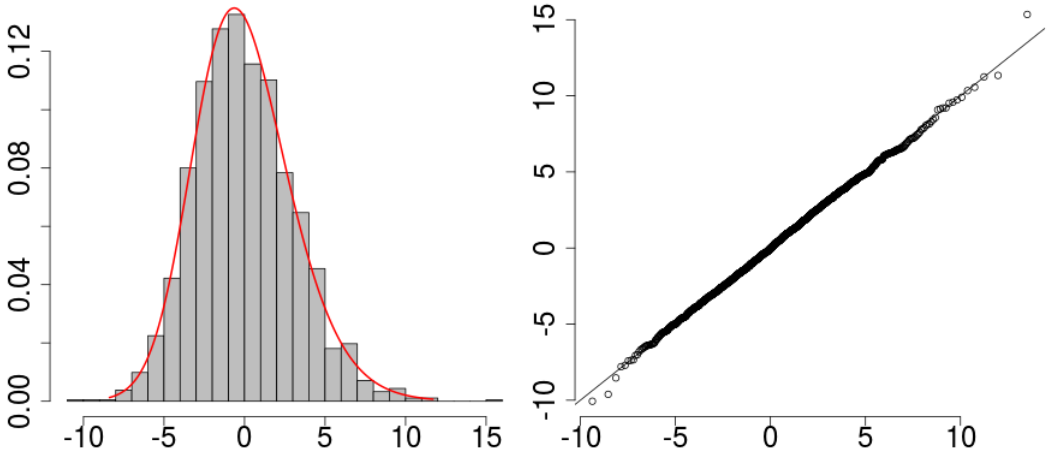


Figure 5.5: The residuals ε_X and the fitted generalised hyperbolic distribution: histogram and q-q plot.

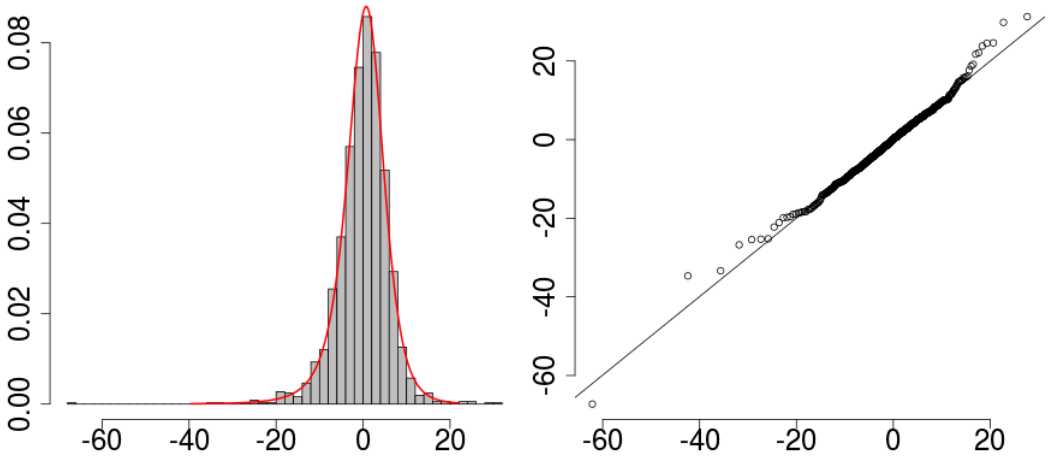


Figure 5.6: The residuals ε_Y and the fitted generalised hyperbolic distribution: histogram and q-q plot.

lines denote 95% confidence intervals computed in bootstrapping procedure with $1,000$ runs. As expected, a shock in the wind energy production has a relatively big impact on electricity prices, while a shock in electricity prices does not significantly influence wind energy generation.

Remark 5.2. The function `VARselect` from the `R` package `vars` run with the maximum lag 10 suggests that the optimal order of the VAR model Eq. (5.1) is:

- 9, according to the Akaike information criterion;
- 6, according to the Hannan-Quinn criterion;
- 3, according to the Schwarz criterion;
- 9, according to the Akaike forecast prediction error.

However, due to the trade-off between the model fit and parsimony, we stick to the lower order. Alternatively, we could consider a VARMA(p,q) model.

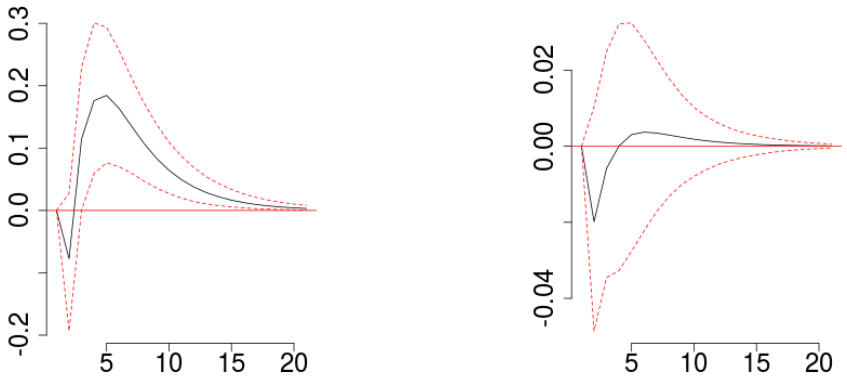


Figure 5.7: Left: response of Y on the impulse from X at time 0. Right: response of X on the impulse from Y at time 0. The red dotted lines denote 95% confidence intervals computed in bootstrapping procedure with $1,000$ runs.

5.2 Value at risk

In this section we propose a way of assessing the model which focuses on its tails, the quality of great importance for practitioners. For simplicity let us use the economic term "profit" to denote $S(t)$, $WD(t)$ or $S(t)WD(t)$. We consider a standard risk measure, value at risk at level $\alpha \in (0,1)$, denoted VaR_α and defined as the $(1-\alpha)$-th quantile of the distribution of interest. Intuitively, the value of risk describes the maximum loss at a given confidence level. For example, let us imagine a contract with a one-month 95% value at risk equal $1,000$ EUR and the expected shortfall $2,000$ EUR. After a month the losses will be bigger than $1,000$ EUR with probability 5% and in that case their expected value will equal $2,000$ EUR.

We estimate a year-long time series of day-ahead values at risk of electricity spot prices $S(t)$, wind energy production $WD(t)$ and their product $S(t)WD(t)$, relevant for contracts such as the one described in Subsection 5.2.2. The product of these variables also serves as a simple alternative to the notion of bivariate quantiles introduced by [Chaudhuri, 1996]. We assume the model Eq. (5.1), calibrated on the 2011–2015 data, and use 2016 data in our computations. We compute values at risk with two levels α – 95% and 99% –

using the Monte Carlo method with $n_{\text{sim}} = 1,000$ simulations.
We start on 31.12.2015 and simulate $X(t)$ and $Y(t)$ for 01.01.2016, using the data from 30.12.2015 and 31.12.2015, since the model requires the data from past two days. We add seasonalities to obtain $S(t)$ and $WD(t)$, the latter after squaring. We compute the value at risk for 01.01.2016. We proceed to compute the value at risk for 02.01.2016 using the data from 31.12.2015 and 01.01.2016, and continue until we obtain $n_{\text{sim}} = 1,000$ time series of day-ahead values at risk for each day of 2016, and each simulation.

5.2.0.1 Backtesting

A good model of value at risk ensures that the true profit (based on the actual data) is lower than the value at risk about $(1 - \alpha)\%$ times. In particular we want to ensure that the value at risk is not set too high, which would underestimate the true risk. We use a standard method of testing a model based on past data: backtesting. For example, the Basel Committee on Banking Supervision uses a traffic light approach to backtesting value at risk ([Bank for International Settlements, 1996]). In this section we consider a simplified version of the procedure required by this regulatory body.
Figure 5.8, Figure 5.9 and Figure 5.10 show that at the first glance one-day-ahead values at risk based on the model Eq. (5.1) behave as expected. To test this idea more rigorously, we can treat each day of 2016 as a Bernoulli trial with the probability of "success" (true profit lower than the value at risk) equal to $1 - \alpha$. This allows us to conduct hypothesis tests on 5% confidence level with three alternative hypotheses: $p \neq p_0$, $p < p_0$ and $p > p_0$, where p_0 denotes the probability of success under the null hypothesis (i.e. $1 - \alpha$), while p is the true probability of success. In other words, p-values bigger than 0.05 indicate a correct model for value at risk, while smaller ones suggest model modifications. Most importantly, we require to fail to reject the latter hypothesis; the alternative would mean that our value at risk is not conservative enough.
In Table 5.5 we present the results of tests conducted with the function `binom.test` from the `R` package `stats` ([R Core Team, 2018]). According to two-sided tests we fail to reject all models apart from the $VaR_{95\%}(WD(t)S(t))$.

However, as we infer from the results of the test with the alternative hypothesis $p > p_0$, the value at risk was in this case set too low, which does not cause as big of a problem as setting it too high.

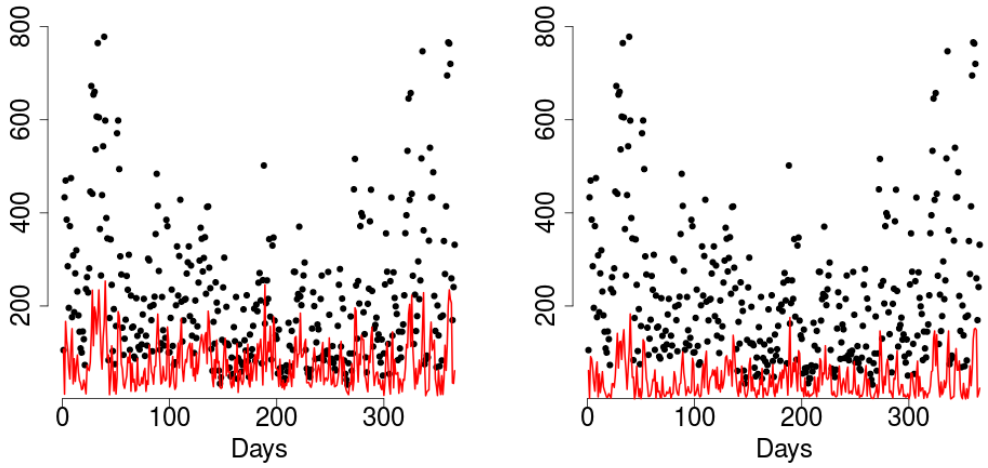


Figure 5.8: Day-ahead values at risk (red line) and true wind energy production forecasts data (dots) for 2016 in GWh.

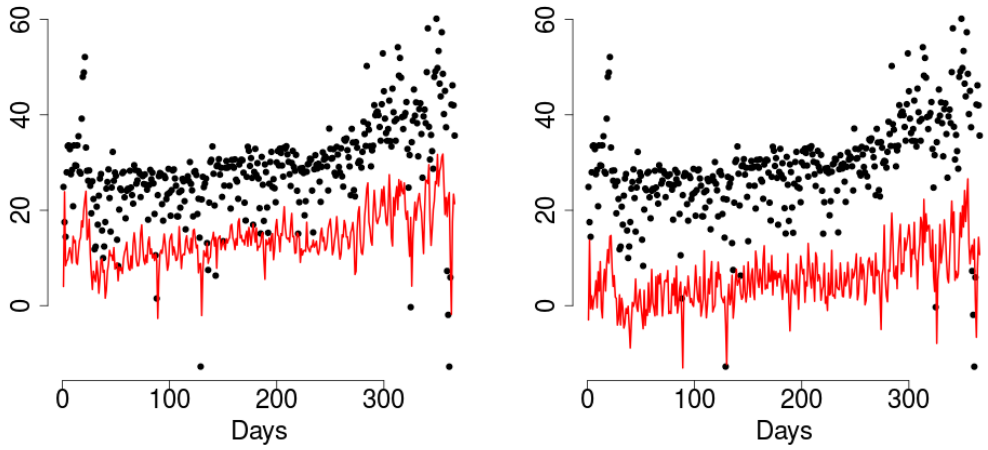


Figure 5.9: Day-ahead values at risk (red line) and true electricity spot prices data (dots) for 2016 in EUR/MWh.

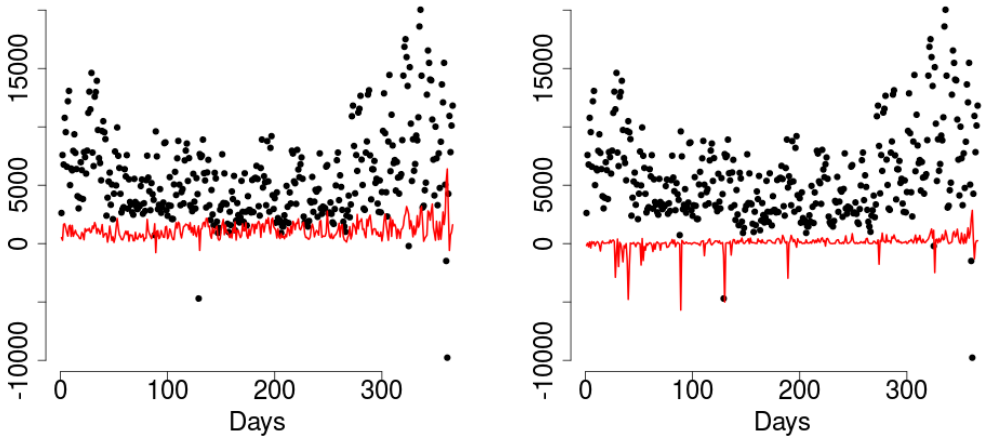


Figure 5.10: Day-ahead values at risk (red line) and true product of wind energy production forecasts and electricity spot prices data (dots) for 2016 in GWh.

variable	$1-\alpha$	exceptions	p-value binomial test with H_A		
			$p \neq p_0$	$p < p_0$	$p > p_0$
$WD(t)$	0.05	20	0.63	0.71	0.37
$WD(t)$	0.01	2	0.6	0.29	0.88
$S(t)$	0.05	14	0.4	0.18	0.88
$S(t)$	0.01	5	0.42	0.84	0.3
$WD(t)S(t)$	0.05	6	0.0016	0.00068	1
$WD(t)S(t)$	0.01	4	0.79	0.7	0.5

Table 5.5: Results of binomial tests of values at risk VaR_α with three alternative hypotheses: true probability of "success" (or exception) different, smaller or greater than the hypothesised one. The column "exceptions" contains the number of instances when the true data was lower than the value at risk.

5.2.1 The joint model versus two separate models

It would be interesting to compare the joint model of wind energy production and electricity prices to two separate models of these variables. As a natural modification of the model Eq. (5.1) we propose two ARMA(p,q) models (Section C.1) fitted to X and Y separately, assuming the independence of variables. To this end we use the function `auto.arima` from the `R` package `forecast`, which uses the stepwise selection of ARMA(p,q) order ([Hyndman and Khandakar, 2008]). It suggests to model X with ARMA(1,1) model with coefficients $\phi_1 = 0.52$ and $\theta_1 = 0.21$, while for Y it chooses ARMA(2,1) with $\phi_1 = 1.35$, $\phi_2 = -0.41$ and $\theta_1 = -0.77$. As in the case of VAR(2) we fit the optimal univariate generalised hyperbolic distributions to residuals of both models: asymmetric variance-gamma to the residuals of X and asymmetric t-Student's to the residuals of Y. In Figure 5.11 and Figure 5.12 we see that both distributions fit well.

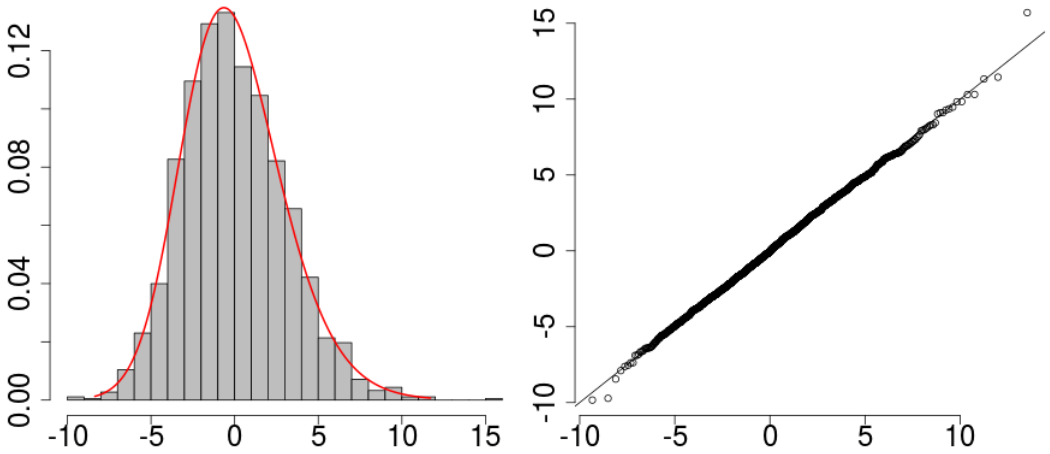


Figure 5.11: The residuals of the ARMA(1,1) model fitted to X and the fitted generalised hyperbolic distribution: histogram and q-q plot.

We recall that the main motivation behind our choice of the joint model Eq. (5.1) instead of two independent models was the ability to capture the relationship between wind energy production and electricity prices. As a proxy

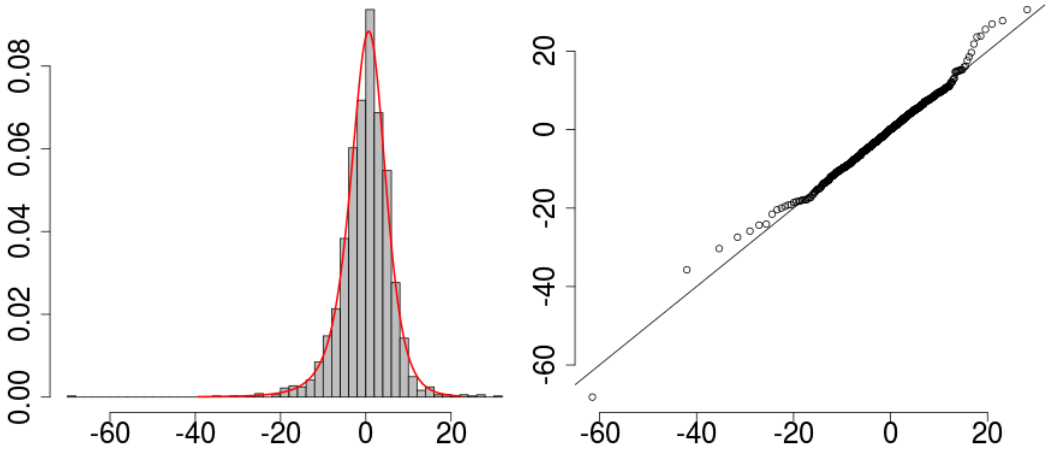


Figure 5.12: The residuals of the ARMA(2,1) model fitted to$Y$ and the fitted generalised hyperbolic distribution: histogram and q-q plot.

to evaluate this feature we use the product of both variables $PR := S \cdot WD$. We simulate 5 years of data from both models, add seasonality and multiply the results to obtain $PR^{\text{VAR}} := S^{\text{VAR}} \cdot WD^{\text{VAR}}$ and $PR^{\text{ARMA}} := S^{\text{ARMA}} \cdot WD^{\text{ARMA}}$. In Figure 5.13 we can already see that the joint model can replicate the true density of the training data better than two ARMA models. The Euclidean distance between true and simulated density confirms this intuition: it is higher in case of independent models. The joint model outperforms independent models especially in higher quantiles, as we observe in Figure 5.14. Therefore while independent models can be suitable to model both variables, we strongly suggest using the joint model to describe the dependence between wind energy production and electricity prices.

5.2.2 Application

In this chapter we proposed a model able to reproduce the dependence between wind energy production and electricity prices. Now we would like to study a contract which requires a way to accurately simulate the product of these variables. Following [Deschatre and Veraart, 2017], we apply the model

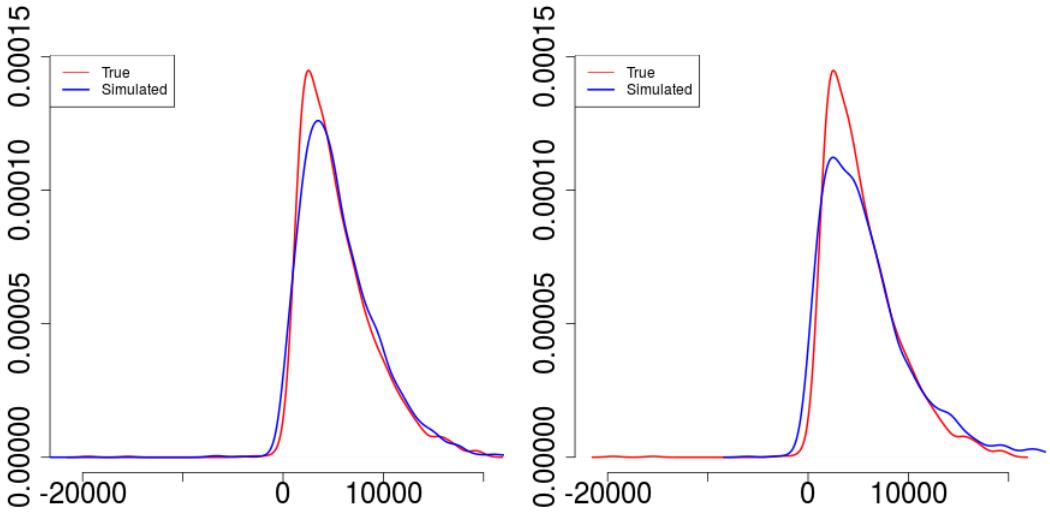


Figure 5.13: True (red) and simulated (blue) densities of the product of wind energy generation and electricity prices, based on the model Eq. (5.1) (left) and two independent ARMA (right) models.

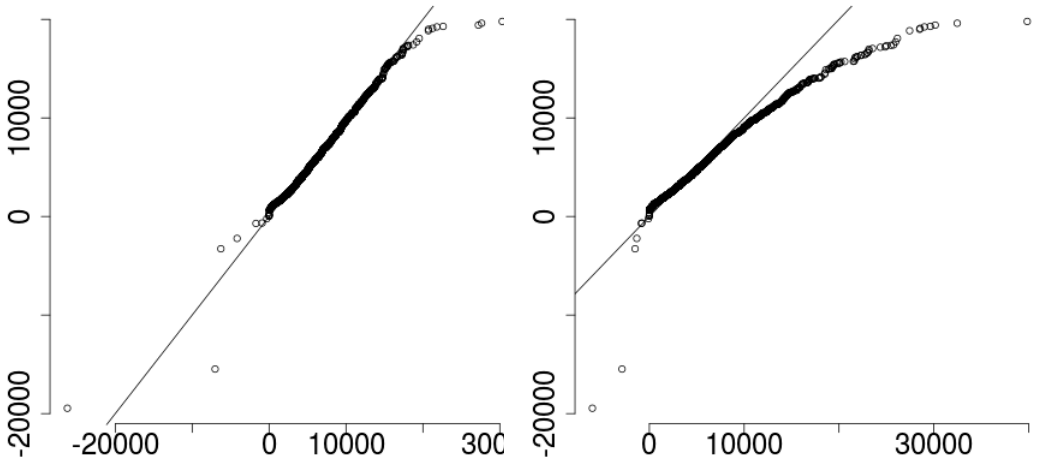


Figure 5.14: The q-q plot of the simulated (horizontal axis) versus true (vertical axis) product of wind energy generation and electricity prices, based on the model Eq. (5.1) (left) and two independent ARMA (right) models.

Eq. (5.1) to quantify the risk faced by electricity distributors. A distributor agrees to pay a fixed price K for the electricity from a wind farm responsible for $Q\%$ of the total wind energy production in a given market. Over a time period from 0 to T the income will equal

$$P = Q \int_0^T (S(t) - K) WD(t) dt.$$

We can study the impact of extreme price values $S(t)$ on the distributor's total income P. We consider two standard risk measures: value at risk (defined in Section 5.2) and the expected shortfall at level $\alpha \in (0, 1)$, denoted $ES_\alpha(P)$ and defined by $\mathbb{E}[P|P \leq VaR_\alpha(P)]$. Intuitively, the expected shortfall tells us about the expected loss in worst case scenarios.

In the following we assume the model Eq. (5.1). We set $Q = 0.01$, $K = 20$ EUR/MWh and the maturity T equal to 366 days (the whole calendar year 2016, which follows the available data set). We start by simulating two time series of length 366: $X(t)$ and $Y(t)$, and adding the seasonality to obtain $S(t)$ and $WD(t)$, the latter after squaring. Further we compute $Q(S(t) - K)WD(t)$ and approximate P by a sum over all points in the simulated $Q(S(t) - K)WD(t)$. We repeat the procedure $n_{\text{sim}} = 1,000$ times to compute VaR and ES. Finally, we use $n_{\text{boot}} = 100,000$ bootstrap samples to compute the confidence intervals of the expected value of P as well as the value at risk and the expected shortfall, all in thousands EUR. Table 5.6 presents 95% confidence intervals for different risk measures with strike price $K = 20$ EUR/MWh. We learn that this particular contract on average is profitable for the distributor.

Risk measure	95% CI		CI length
Expectation	788.74	864.61	75.87
$VaR_{95\%}(P)$	-1175.96	-1065.37	110.59
$ES_{95\%}(P)$	-8204.39	-6710.64	1493.75
$VaR_{99\%}(P)$	-5648.29	-4723.55	924.74
$ES_{99\%}(P)$	-34132.14	-25818.15	8313.99

Table 5.6: Confidence intervals of the expectation and risk measures of the profit P in thousands EUR, with the strike price $K = 20$ EUR/MWh.

5.3 Summary and outlook

In this chapter we jointly modelled wind energy production and electricity spot prices. We based our model on a bivariate vector autoregressive process with correlated generalised hyperbolic residuals. The model captures very well the correlation structure; in particular, it outperforms univariate models when applied to the product of two variables of interest, a value relevant for many applications. When validating the model we focused on low quantiles: we assessed the value at risk, a measure of great importance for practitioners. The backtesting procedure indicated the the model accurately estimated the risk related to wind energy production and electricity prices as well as their product. Finally, we computed the value at risk and expected shortfall of a contract relating these variables and their product.

We believe that this joint modelling approach might be improved even further if one could get better marginal fits while still capturing the correlation structure. We suggest exploring the copula approach ([Czado, 2019]). One could also include other wind-related variables such as the wind penetration index or residual demand (Chapter 3). Finally, one could extend the vector autoregressive model by the moving average component which might improve the results even further.

Part II

Gaussian moving average processes applied to wind speed modelling and forecasting

6

Introduction

6.1 Motivation

As we observed in Chapter 4, wind energy production forecasts highly influence electricity prices. However, accurate predictions of wind energy generation still pose a serious challenge. In this part of the thesis we attempt to fill this gap by modelling wind speed in one location. We aim to produce reliable models and short-term forecasts of wind speed, which for a specific turbine can be converted into wind energy production through a power curve ([Zhu and Genton, 2012]).

Such local (at a wind farm level) models and forecasts are of particular interest to practitioners. For example, the wind farm's owner needs to predict her or his production levels to prevent energy shortages and unnecessary losses. Furthermore, all members of the European Union must release Guarantees of Origin (GoOs) which certify that a given MWh of energy comes from renewable sources ([European Commission, 2009]). Therefore more and more supply contracts promise the consumers to provide an agreed percentage of energy from local wind farms, which further increases the importance of accurate wind energy predictions.

We use the statistical approach to wind speed forecasting, which involves historical data as opposed to computationally expensive Numerical Weather Prediction (NWP) models. According to [Giebel, 2003], statistical models can outperform physical models for predictions up to six hours ahead, therefore we focus on this time frame.
We propose a temporal model for wind speed in one location corresponding to a small group of wind turbines. Although this approach excludes the information about high spatial correlation, typical for weather variables, we believe that we can alleviate this problem by including the wind direction data in our analysis. By reducing the spatial complexity, we can use high frequency data (5-minute intervals) to better model strong temporal correlations reaching even over 24 hours ([Hering and Genton, 2010]).

6.2 Wind speed forecasting in the literature

Out of the vast literature about wind speed modelling, a few models have caught our attention. [Gneiting et al., 2006] proposed a spatio-temporal model with westerly and easterly regimes pronounced in Oregon and Washington, US. [Hering and Genton, 2010] argued that two regimes might not suffice, so they replaced directional regimes by trigonometric functions of wind speed directions as covariates. Finally, [Benth and Benth, 2010] used more standard methods: they applied the Box-Cox transformation to the right-skewed wind speed data and proceeded to model it as an autoregressive moving average process (cf. Section C.1).
We must also mention a class of approaches which has recently gained a lot of popularity in the field: machine learning. For example, [Ak et al., 2016] applied two different machine learning methods to short-term forecasts of wind speed in Saskatchewan (Canada), while [Sergio and Ludermir, 2015] studied the performance of deep learning approaches used to forecast wind speeds in the north-eastern part of Brazil. Recently researchers from Google's DeepMind successfully trained a neural network on historical weather and wind turbine data to forecast wind turbine output 36 hours ahead [Witherspoon and Fadrhonc, 2019].

Machine learning approaches are getting increasingly better results, especially given growing capacities of modern computers. However, we are interested in exploring a different method for a variety of reasons. Firstly, we would like to be able to forecast wind speeds in a relatively short time without the access to supercomputers. Secondly, we want to explore models given by explicit formulae to assess their mathematical properties rigorously, instead of using a black-box model. Thirdly, machine learning approaches are usually hard to interpret, as opposed to more traditional models.
Surprisingly few authors have considered forecasting horizontal wind velocity components. [Hering and Genton, 2010] extended the regime-switching framework suggested by [Gneiting et al., 2006] and included wind direction as a circular variable. [Pinson, 2012], [Schuhen et al., 2012] and [Lang et al., 2019] forecasted zonal and meridional components of wind speed with bivariate Gaussian processes. [Sloughter et al., 2013] produced probabilistic forecasts of wind vectors using Bayesian methodology. Finally, [Monahan, 2013] pointed out that these bivariate approaches have multiple advantages over modelling the wind speed directly: horizontal components include information about both speed and direction; their joint distribution resembles bivariate Gaussian more than the joint distribution of speed and direction; dynamical models usually require the input of vector components instead of speed and direction. Our work further explores this two-component approach to wind speed forecasting.

6.3 Outline

We structure this part of the thesis as follows. We start by describing the data sets as well as the procedures of deseasonalisation and splitting wind velocity into horizontal components (Chapter 7). In Chapter 8 we move on to describing the one-dimensional Gaussian moving average model of a horizontal component. We discuss the model's properties (Section 8.1) as well as the estimation (Section 8.2) and forecasting methods (Section 8.3). In Chapter 9 we extend this model to two dimensions. We describe the model (Chapter 9), discuss the estimation procedures (Section 9.2), look at

forecasting (Section 9.3) and simulating from the model (Section 9.4). In Section 9.5 we compare our results to a simpler approach of forecasting wind speed directly.

7

The data

7.1 Data description

Our data set consists of wind speeds $\tilde{S}(t)$ in $\mathrm{m\,s^{-1}}$ and directions $\tilde{R}(t)$ in degrees from the North measured by the Institut Pierre Simon Laplace (IPSL) Atmospheric Research Observatory, Site Instrumental de Recherche par Télédétection Atmosphérique (SIRTA); for convenience we use the tilde to denote raw data, reserving bare letters for deseasonalised variables. We work with 1-minute averages of 10-second data from a 10m mast located on a laboratory roof platform at 15m above ground level, near Paris (48.7°N, 2.2°E), created by Christophe Boitel ([Haeffelin et al., 2005]). The available data set starts on 01.01.2007 and finishes on 12.01.2017. In this chapter we analyse a subset of this data encompassing 01.03.2010–31.05.2010, presented in Figure 7.1, as an example of all time periods available to us. We notice that northerly winds slightly dominate in the specified time period (this dominance is much more pronounced in other seasons of 2010).
From wind speeds $\tilde{S}(t)$ and wind directions $\tilde{R}(t)$ we obtain horizontal compo-

nents of wind speed using the following relations:

$$\tilde{X}(t) = \tilde{S}(t) \cos\left(\frac{\tilde{R}(t)\pi}{180^\circ}\right),$$
$$\tilde{Y}(t) = \tilde{S}(t) \sin\left(\frac{\tilde{R}(t)\pi}{180^\circ}\right).$$

We notice that $\tilde{X}$ corresponds to the North–South, while $\tilde{Y}$ to the East–West component of the wind. In particular:

- Northerly winds ($\tilde{R}(t) = 0^\circ$) result in $\tilde{X} = \tilde{S}$ and $\tilde{Y} = 0$.
- Easterly winds ($\tilde{R}(t) = 90^\circ$) result in $\tilde{X} = 0$ and $\tilde{Y} = \tilde{S}$.
- Southerly winds ($\tilde{R}(t) = 180^\circ$) result in $\tilde{X} = -\tilde{S}$ and $\tilde{Y} = 0$.
- Westerly winds ($\tilde{R}(t) = 270^\circ$) result in $\tilde{X} = 0$ and $\tilde{Y} = -\tilde{S}$.

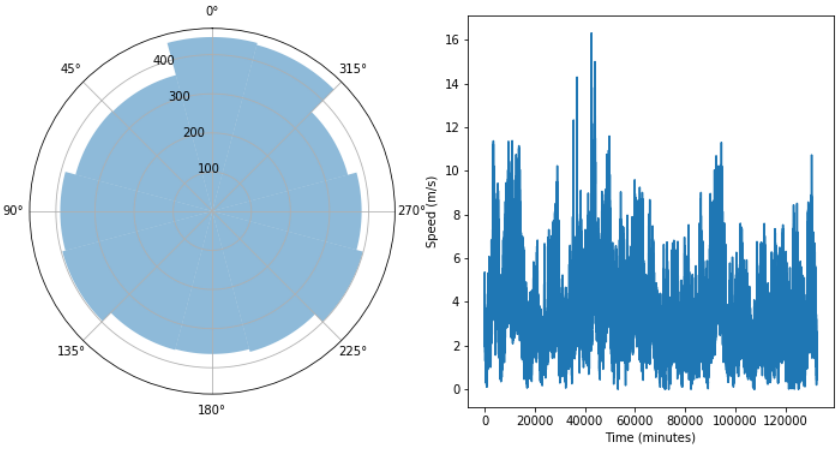


Figure 7.1: Histogram of wind directions measured from the North (0° corresponds to northerly winds, 90° to easterly etc.) and time series of wind speeds, 01.03.2010–31.05.2010.

7.2 Smoothing and seasonality

Like most climatological time series, our data set displays strong seasonality, both on daily and yearly scales. In this section we first describe the Savitzky-Golay filter and show how we apply it to smooth out the data set.

7.2.1 Savitzky-Golay filter

For $i = \ldots, -2, -1, 0, 1, 2, \ldots$ and $\Delta > 0$ let $f_i := f(\Delta i)$ denote a series of equally spaced data. We can smooth f_i out by applying a digital filter ([Press et al., 1992]). We replace each value of f_i by g_i, a linear combination of itself and its neighbours, n_L preceding and n_R following, which means that

$$g_i = \sum_{n=-n_L}^{n_R} c_n f_{i+n}.$$

The simplest digital filter, the so-called moving window average filter, assumes that $c_n = \frac{1}{n_L+n_R+1}$. This filter preserves the zeroth and first moments, but distorts higher moments of the function. The Savitzky-Golay filter generalises this idea by finding such coefficients c_n that higher moments get preserved as well. In other words, instead of replacing the underlying function by a constant, we use the least squares method to fit polynomials of higher degrees. Figure 7.2 shows an example of smoothing exponential data with Gaussian noise using a Savitzky-Golay filter with the length of moving window 31 and polynomial order 4.

7.2.2 Smoothing and deseasonalisation procedure

We remove the seasonal patterns from $\tilde{S}$, $\tilde{X}$ and $\tilde{Y}$ as follows.

1. We introduce 2880 variables: means and standard deviations for each minute of the day.

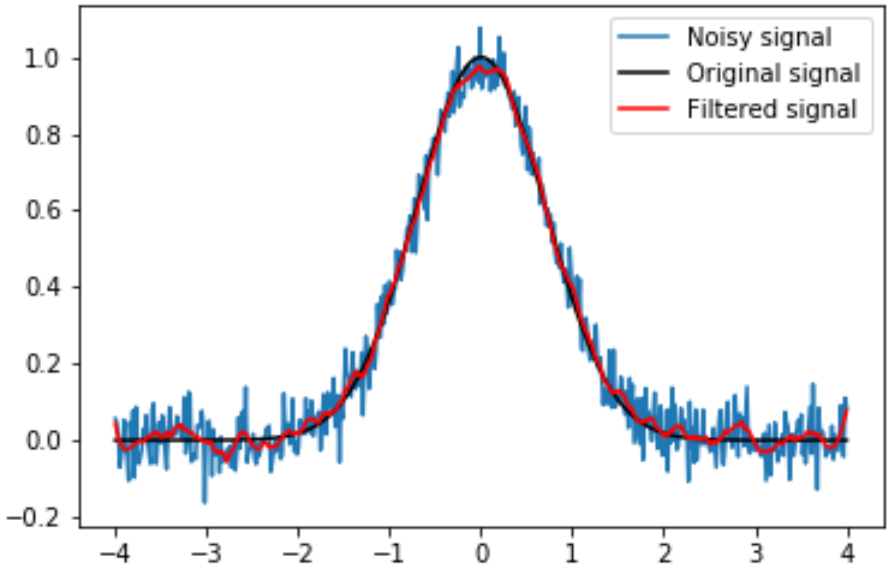


Figure 7.2: Results of Savitzky-Golay filter with the length of moving window 31 and polynomial order 4. The data $y(t) = e^{-t^2}$ was perturbed with a noise $\varepsilon \sim N(0, 0.05)$.

2. We smooth out the resulting time series means and standard deviations using the Savitzky-Golay filter with the window length 121 and polynomial order 3.

3. We subtract the smoothed means from the appropriate hours of original time series and divide the results by appropriate smoothed standard deviations.

We acknowledge that 2880 might seem like a large number of new variables. However, a coarser time scale (for example 48 variables: means and standard deviations for each hour of the day) fails to remove the seasonality from the data.
The means and standard deviations computed in the first deseasonalisation step are contaminated with a random noise. Therefore we smooth out these time series with a low-pass Savitzky-Golay filter ([Savitzky and Golay, 1964]). While simple moving average filtering techniques tend to reduce the noise more effectively, the Savitzky-Golay filter better preserves the original shape

and features of the signal. This algorithm approximates the desired function within the moving window by a non-constant polynomial fitted by the method of least squares.

Because we are interested in short-term forecasting, we do not want to remove the short-term variations from the original data set. Therefore we apply the filter only to seasonal dummy variables as opposed to the raw data.

Figure 7.3 presents the original and filtered daily means and daily standard deviations of wind speed and its horizontal components. We are satisfied with the stationarity of intraday means of deseasonalised horizontal components. While some non-stationary patterns persist, mostly in standard deviations, they do not pose a significant problem, as in our research we focus on very short-term forecasts, for which non-stationarities disappear. One can argue that in this particular application we could avoid such a complex smoothing procedure. However, we aim to make our research as general and widely applicable as possible, so we decided to use this flexible framework.

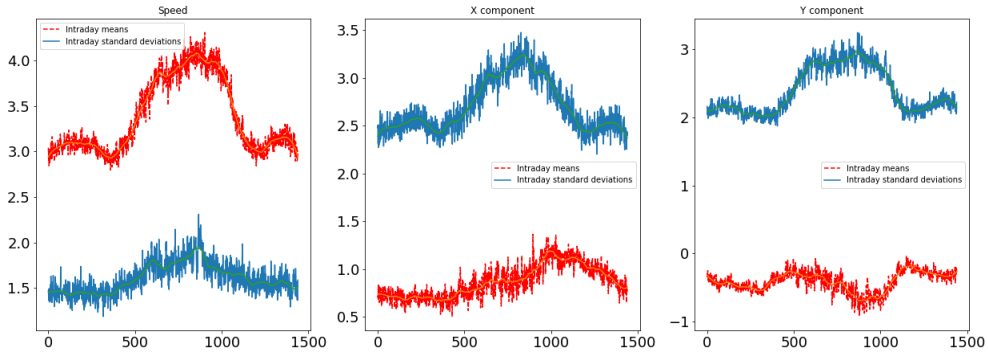


Figure 7.3: Intraday means (red) and standard deviations (blue) of wind speed S and its horizontal components X and Y, 01.03.2010–31.05.2010, original and smoothed by Savitzky-Golay filter with the length of moving window 121 and polynomial order 3.

7.3 Data analysis

We denote the deseasonalised time series $\tilde{S}$, $\tilde{X}$ and $\tilde{Y}$ by S_{des}, X and Y, respectively. Because our deseasonalising process is non-linear and results in negative values of S_{des}, we additionally define

$$S = \sqrt{X^2 + Y^2}$$

to get meaningful speed values. As we can see in Figure 7.4, while the smoothed wind speed clearly deviates from Gaussian distribution, the Gaussian law reasonably approximates the X and Y projections.

The correlation coefficient between X and Y equals 0.40, similarly Figure 7.5 indicates a mild positive correlation between the deseasonalised components. The correlation coefficient between the original components $\tilde{X}$ and $\tilde{Y}$ is almost the same as between X and Y: 0.39. Non-deseasonalised speed and direction are almost uncorrelated, with the coefficient -0.15. Such low dependence seems to be specific to our data set, not a universal feature of wind speeds and directions.

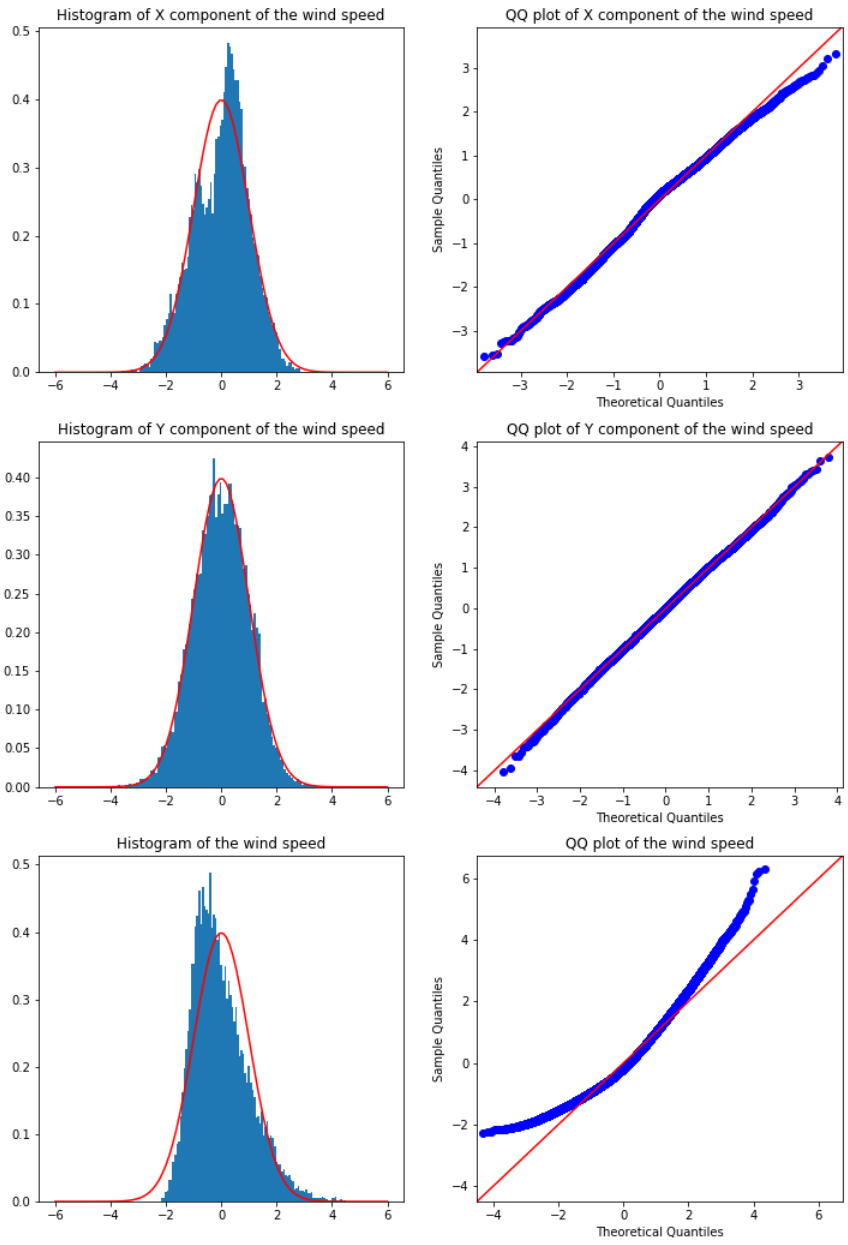


Figure 7.4: Distribution and qq-plots of two smoothed components X and Y and smoothed wind speed S_{des}, 01.03.2010–31.05.2010. Data sampled every 10 minutes.

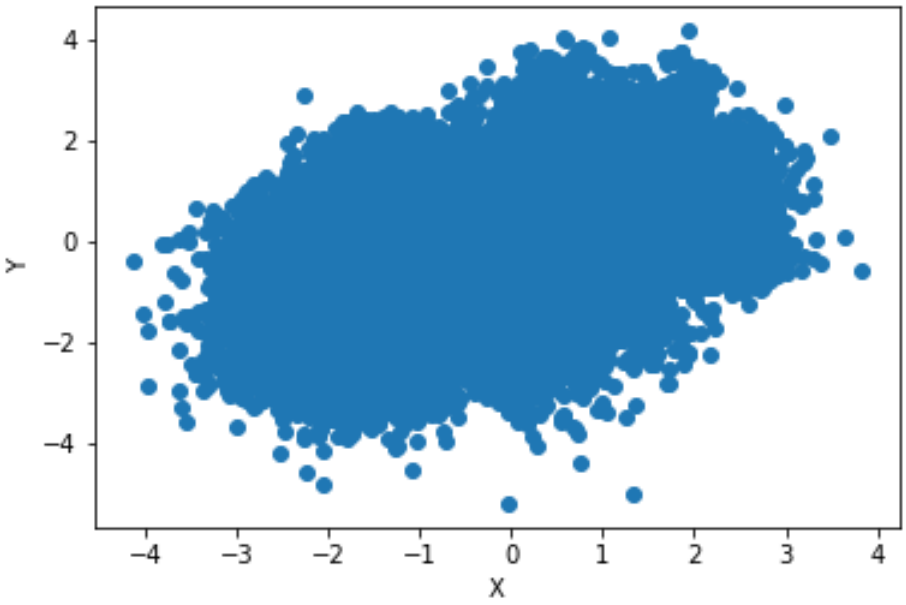


Figure 7.5: The scatter plot of deseasonalised components X and Y, 01.03.2010–31.05.2010.

8

Modelling horizontal components

In this chapter we separately model both horizontal components of wind speed obtained in Chapter 7. The models are based on Gaussian moving average processes with a novel kernel function which generalises gamma and power-law kernels, called the *triple-scale* kernel. We explore two estimation methods: the generalised method of moments (Subsection 8.2.1.1) and the regression-based method (Subsection 8.2.1.2). Based on the estimation results, in Section 8.3 we compute short-term forecasts of wind speed and analyse their accuracy.

8.1 One-dimensional model

The Gaussian moving average process belongs to a flexible class of Gaussian processes with a rich correlation structure, which makes it suitable for modelling horizontal wind speed components. We define it via the moving-average representation

$$Z(t) = \int_{-\infty}^{t} g(t-s)dW(s)\,, \tag{8.1}$$

where $t \in \mathbb{R}$, W denotes a standard Brownian motion and g a square integrable kernel function. Any stationary, non-deterministic, centred and mean square continuous-time Gaussian process admits a representation of this form [Hida and Hitsuda, 1993, paragraph III.4]. A moving average Gaussian process is stationary and for $h \in \mathbb{R}$ its covariance function has the form

$$c(t, t+h) = \int_0^\infty g(t)g(t+h)dt := c(h).$$

For $h \in \mathbb{R}$ we also introduce the *variogram* function

$$v(h) = \mathbb{E}[(Z(t+h) - Z(t))^2] = 2(c(0) - c(h)).$$

Gaussian moving average processes belong to the class of Lévy semistationary processes ([Barndorff-Nielsen and Shephard, 2012]). Examples of kernels widely used in the literature (e.g. [Barndorff-Nielsen et al., 2013] or [Bennedsen et al., 2017]) include:

- the exponential kernel (Ornstein-Uhlenbeck process): $g_0(t) = \exp\left(-\frac{t}{\Delta_2}\right)$ for $\Delta_2 > 0$;
- the gamma kernel: $g_1(t) = t^\alpha \exp\left(-\frac{t}{\Delta_2}\right)$ for $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$ and $\Delta_2 > 0$;
- the power-law kernel: $g_2(t) = t^\alpha (1+t)^{\beta-\alpha}$ for $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$ and $\beta < -\frac{1}{2}$,

where $t > 0$. In Figure 8.1 we can observe three time scales present in empirical variograms of horizontal components of wind speed (especially Y). Therefore we suggest modelling the horizontal components with the following generalisation of gamma and power-law kernels:

$$\begin{aligned} g_X(t) &= \sigma_X t^{\alpha_X} \left(1 + \frac{t}{\Delta_{1,X}}\right)^{\beta_X} \exp\left(-\frac{t}{\Delta_{2,X}}\right), \\ g_Y(t) &= \sigma_Y t^{\alpha_Y} \left(1 + \frac{t}{\Delta_{1,Y}}\right)^{\beta_Y} \exp\left(-\frac{t}{\Delta_{2,Y}}\right), \end{aligned} \tag{8.2}$$

where $\theta = (\sigma, \alpha, \beta, \Delta_1, \Delta_2)$, $\alpha > -\frac{1}{2}$, $\beta \in \mathbb{R}$, $\Delta_1 > 0$, $\Delta_2 > 0$ and $\sigma > 0$. Such triple-scale kernels reproduce three time scales in variograms of our wind time series better than gamma or power-law kernels, recovered by setting $\beta = 0$ or $\Delta_2 = \infty$, respectively (for details see Subsection 8.3.2.1).

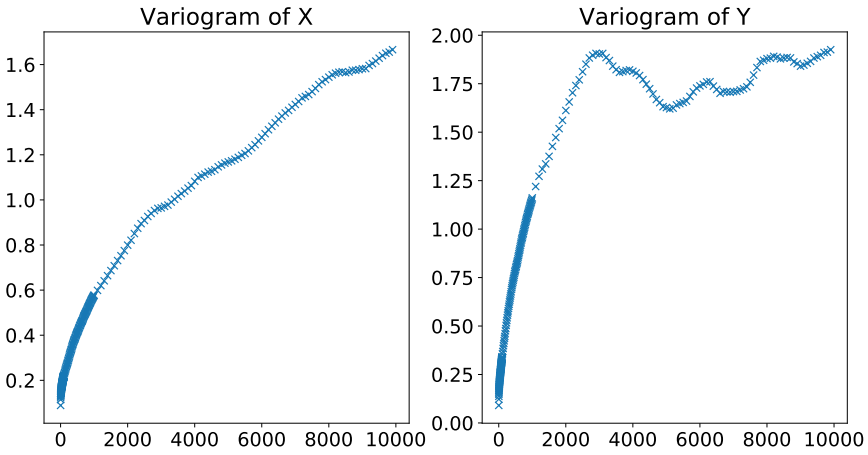


Figure 8.1: Variograms of deseasonalised components X and Y, 01.03.2010–31.05.2010. Lags correspond to minutes.

8.2 Estimation

In this section we introduce two estimation methods of the model presented in Section 8.1: the generalised method of moments (Subsection 8.2.1.1) and a regression-based method (Subsection 8.2.1.2).

8.2.1 Methods

8.2.1.1 The generalised method of moments

For proofs and further details we refer the reader to [Matyas et al., 1999] or [Hall, 1996]. Suppose we want to use an N-dimensional vector of observations Z to estimate a parameter vector θ with a true value θ_0. Let $f^{\mathrm{GMM}}(Z, \theta)$ be a continuous vector function of θ such that $\mathbb{E}\left[f_t^{\mathrm{GMM}}(Z_t, \theta_0)\right] = 0$ for all

$t = 1, \ldots, N$. We define the criterion function

$$Q(\theta) = \left(\bar{f}^{\mathrm{GMM}}(Z, \theta)\right)^T W \bar{f}^{\mathrm{GMM}}(Z, \theta),$$

where the superscript T denotes matrix transpose, W is a positive definite GMM weight matrix and $\bar{f}^{\mathrm{GMM}}$ the empirical estimate of the expected value of f^{GMM}. We define the GMM estimator as

$$\arg\min_{\theta} Q(\theta),$$

which we can rewrite as

$$\arg\min_{\theta} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} W_{mn} \bar{f}_m^{\mathrm{GMM}}(Z_m, \theta) \bar{f}_n^{\mathrm{GMM}}(Z_n, \theta).$$

Such an estimator is consistent and asymptotically normal ([Hall, 1996, Chapter 3]). The optimal weight matrix resulting in the lowest variance of the estimator is given by $W^* = (\Omega^\theta)^{-1}$, where

$$\Omega_{mn}^{\theta} = \mathbb{E}[\left(f^{\mathrm{GMM}}(Z, \theta)\right)^T f^{\mathrm{GMM}}(Z, \theta)].$$

Because the matrix Ω depends on the unknown parameter value θ_0, we use a two-step procedure.

1. We compute the initial estimator $\tilde{\theta}$ with a unit weight matrix.
2. We compute the optimal weight matrix using $\widehat{W}^* = (\Omega^{\tilde{\theta}})^{-1}$ and recompute the estimator $\hat{\theta}$ with the weight matrix $\widehat{W}^*$.

8.2.1.2 The regression-based method

Plots such as Figure 7.1 suggest that wind speeds and their components have a characteristic named by some researchers *antipersistence* or *roughness* ([Bennedsen, 2015]). We quantify the roughness or smoothness of time series data, relative to paths of a Brownian motion, with a fractal index $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Curves with $\alpha < 0$ are rougher, while $\alpha > 0$ smoother than

paths of a Brownian motion characterised by $\alpha = 0$. The most common example of a Gaussian process with a positive fractal index is the fractional Brownian motion (fBm) with Hurst parameter $H \in (0, 1]$ and $\alpha = H - \frac{1}{2}$ ([Bennedsen, 2015]).
Inspired by the approach of [Bennedsen et al., 2017], we use the theory of fractals to estimate the parameters of one-dimensional Gaussian moving average processes. We need to introduce the following concepts.

Definition 8.1. We call a measurable function $L : (0, 1] \to [0, \infty)$ *slowly varying at 0* if for all $t > 0$

$$\lim_{x \to 0} \frac{L(tx)}{L(x)} = 1.$$

Definition 8.2. We say that a function $f(x) = \mathcal{O}(h(x))$ as $x \to \infty$ if and only if

$$\limsup_{x \to \infty} \left| \frac{f(x)}{h(x)} < \infty \right|.$$

Following [Bennedsen et al., 2017] we consider a kernel g for which the following conditions hold.

(C1) There exits a parameter $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$ such that for all $t \in (0, 1]$ $g(t) = t^\alpha L_g(t)$, where $L_g(t)$ is continuously differentiable, slowly varying at 0 and bounded away from 0. There also exists a constant $C > 0$ such that for all $t \in (0, 1]$ the derivative $L_g^{(1)}$ of L_g satisfies $|L_g^{(1)}(t)| \leq C\,(1 + t^{-1})$.

(C2) The function g is continuously differentiable on $(0, \infty)$, so that its derivative $g^{(1)}$ is ultimately monotonic and satisfies $\int_1^\infty g^{(1)}(t)^2 dt < \infty$.

(C3) There exists $\xi > \frac{1}{2}$ such that $g(t) = \mathcal{O}(t^{-\xi})$ as $t \to \infty$.

We have already encountered an example of such kernel: the triple-scale kernel.

Proposition 8.0.1. *For $t > 0$ the kernel function*

$$g(t) = \sigma t^{\alpha} \left(1 + \frac{t}{\Delta_1}\right)^{\beta} \exp\left(-\frac{t}{\Delta_2}\right). \tag{8.3}$$

satisfies conditions (C1)–(C3).

Proof. (C1) We define

$$L_g(t) := \sigma \left(1 + \frac{t}{\Delta_1}\right)^{\beta} \exp\left(-\frac{t}{\Delta_2}\right).$$

This function is clearly continuously differentiable. Since

$$\lim_{x\to 0} \frac{L_g(tx)}{L_g(x)} = \lim_{x\to 0} \frac{\left(1 + \frac{tx}{\Delta_1}\right)^{\beta} \exp\left(-\frac{tx}{\Delta_2}\right)}{\left(1 + \frac{x}{\Delta_1}\right)^{\beta} \exp\left(-\frac{x}{\Delta_2}\right)} = 1,$$

it is also slowly varying at 0.

(C2) For $t > 0$ we compute

$$g^{(1)}(t) = g(t)\left[\frac{\alpha}{t} + \frac{\beta}{\Delta_1 + t} - \frac{1}{\Delta_2}\right],$$
$$g^{(2)}(t) = g(t)\left[\left(\frac{\alpha}{t} + \frac{\beta}{\Delta_1 + t} - \frac{1}{\Delta_2}\right)^2 - \frac{\alpha(\Delta_1 + t)^2 + \beta t^2}{t^2\,(\Delta_1 + t)^2}\right].$$

Therefore

$$\lim_{t\to\infty} g^{(2)}(t) = \frac{1}{\Delta_2^2} \lim_{t\to\infty} g(t) > 0,$$

so the first derivative $g^{(1)}(t)$ is ultimately increasing. Moreover, for $t \geq 1$

$$g^{(1)}(t)^2 = g(t)^2 \left[\frac{\alpha}{t} + \frac{\beta}{\Delta_1 + t} - \frac{1}{\Delta_2}\right]^2 \leq g(t)^2 \left[\alpha + \frac{\beta}{\Delta_1 + 1} - \frac{1}{\Delta_2}\right]^2,$$

so $\int_1^{\infty} g^{(1)}(t)^2 dt < \infty$ as $g(t)$ is square integrable.

(C3) It suffices to notice that $g(t)$ decays exponentially to 0 as $t \to \infty$.

□

Now we are ready to apply the properties of the kernel function to approximate the variogram of the moving-average Gaussian process.

Proposition 8.0.2. *For a small number of lags h we can approximate the variogram $v(h)$ of a moving average Gaussian process*

$$Z(t) = \int_{-\infty}^{t} g(t-s)dW(s)$$

with a square integrable kernel $g(t)$ satisfying conditions (C1), (C2) and (C3) by

$$v(h) \approx L_g(h)^2 h^{2\alpha+1} \frac{\Gamma^2(\alpha+1)}{\Gamma(2\alpha+2)\cos(\pi\alpha)}, \quad h \to 0$$

where

$$(f_1(t) \approx f_2(t), \quad t \to 0) \iff \lim_{t\to 0} \frac{f_1(t)}{f_2(t)} = 1.$$

Proof. Using [Bennedsen et al., 2017, Proposition 2.2] we can approximate

$$v(h) \approx h^{2\alpha+1} \left(\int_0^\infty (s^\alpha - (s+1)^\alpha)^2 ds + \frac{1}{2\alpha+1} \right) L_g(h)^2, \quad h \to 0. \tag{8.4}$$

On the other hand, computing the variance of the fractional Brownian motion $W^H(t)$ with the Mandelbrot-Van Ness representation ([Mandelbrot and Van Ness, 1968]), we get

$$\begin{aligned}\mathbb{E}[(W^H(t))^2] &= \frac{1}{\Gamma^2(H+1/2)} \left(\int_0^\infty (s^{H-1/2} - (s+1)^{H-1/2})^2 ds + \frac{1}{2H} \right) \\ &= \frac{1}{\sin(\pi H)\Gamma(2H+1)}.\end{aligned}$$

We substitute $H = \alpha + \frac{1}{2}$ and transform the equation to obtain

$$\int_0^\infty (s^\alpha - (s+1)^\alpha)^2 ds + \frac{1}{2\alpha+1} = \frac{\Gamma^2(\alpha+1)}{\Gamma(2\alpha+2)\cos(\pi\alpha)}. \tag{8.5}$$

From Eq. (8.4) and Eq. (8.5) we conclude that

$$v(h) \approx L_g(h)^2 h^{2\alpha+1} \frac{\Gamma^2(\alpha+1)}{\Gamma(2\alpha+2)\cos(\pi\alpha)}, \quad h \to 0.$$

□

We apply Proposition 8.0.2 to Eq. (8.3) and obtain

$$v(h) \approx \sigma^2 h^{2\alpha+1} \frac{\Gamma^2(\alpha+1)}{\Gamma(2\alpha+2)\cos(\pi\alpha)},$$

as in this case

$$L_g(h)^2 \approx \sigma^2, \quad h \to 0.$$

Therefore we can regress the logarithm of the variogram function on the logarithm of the lag number

$$\log(v(h)) \approx m \log(h) + b$$

to estimate the parameters α and σ:

$$\hat{\alpha} = \frac{m-1}{2}; \tag{8.6}$$

$$\hat{\sigma} = \frac{\sqrt{\exp(b)\Gamma\left(2\hat{\alpha}+2\right)\cos\left(\pi\hat{\alpha}\right)}}{\Gamma\left(\hat{\alpha}+1\right)}. \tag{8.7}$$

We estimate the remaining parameter β by matching theoretical and empirical variograms using the method of least squares.

8.2.2 Numerical results

We apply both estimation methods described in Section 8.2 to the Gaussian moving average process with the triple-scale kernel.

8.2.2.1 The generalised method of moments

We estimate the parameters θ_X and θ_Y of the horizontal components using the generalised method of moments described in Subsection 8.2.1.1. Let $\hat{Z}$ denote empirical data and $h \in \{1, 5, 10, 20, 30, 60, 120, 180, 360\}$ (minutes). We define

$$f_1^{\mathrm{GMM}}(Z, \theta, h) := (Z(t+h) - Z(t))^2) - (\hat{Z}(t+h) - \hat{Z}(t))^2.$$

Then

$$f^{\mathrm{GMM}}(Z, \theta) = \begin{bmatrix} f_1^{\mathrm{GMM}}(Z, \theta, 1) \\ \vdots \\ f_1^{\mathrm{GMM}}(Z, \theta, 360) \end{bmatrix}.$$

We start the least squares procedure with $(\sigma_X^0, \alpha_X^0, \beta_X^0) = (\sigma_Y^0, \alpha_Y^0, \beta_Y^0) = (0.1, 0.1, 0.1)$. We fit the model to the first 180 lags of the variogram function. We use the function `minimize` from Python's library `scipy.optimize` ([Virtanen et al., 2019]) with the limited-memory Broyden-Fletcher-Goldfarb-Shanno method (L-BFGS-B, [Zhu et al., 1997]) and bounds $(\sigma_X, \alpha_X, \beta_X) \in (0.0, \infty) \times (-0.5, \infty) \times (-\infty, \infty)$ and $(\sigma_Y, \alpha_Y, \beta_Y) \in (0.0, \infty) \times (-0.5, \infty) \times (-\infty, \infty)$. To approximate the integrals numerically, we use the function `quad` from Python's library `scipy.integrate` ([Virtanen et al., 2019], based on 15-point Gauss-Kronrod quadrature. For integrals with at least one infinite limit, the software computes a Fourier integral.
We use only the first step of the GMM procedure, as the second step is computationally expensive and results in numerical instabilities. Figure 8.2 presents empirical and fitted variograms for both horizontal components, for 01.03.10–31.05.10.

8.2.2.2 The regression-based method

We estimate the parameters θ_X and θ_Y of the horizontal components using the regression-based method described in Subsection 8.2.1.2. To regress the logarithm of the variogram function on the logarithm of the lag number, we use the function `linregress` from Python's library `scipy.stats` ([Virtanen

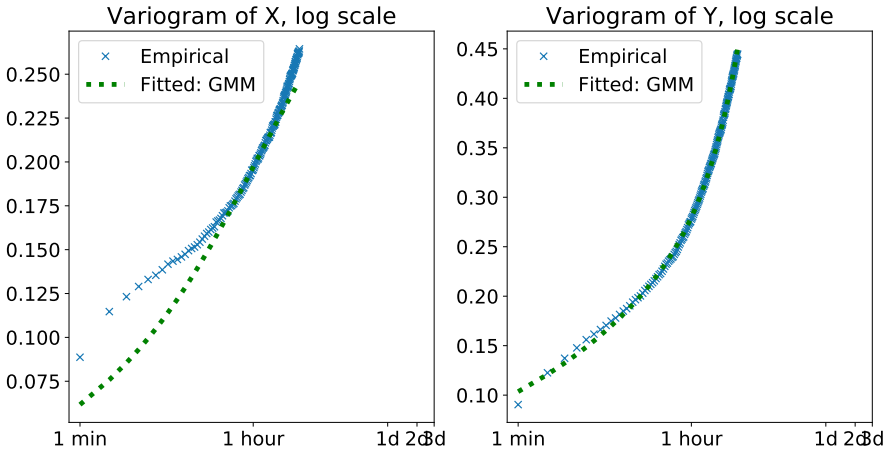


Figure 8.2: Variograms of horizontal components fitted with the generalised method of moments (GMM), using first 180 lags. Training set: 01.03.10–31.05.10.

et al., 2019]). The remaining parameter β we fit similarly to the generalised method of moments case, using the function `minimize` from Python's library `scipy.optimize` ([Virtanen et al., 2019]) with the Broyden-Fletcher-Goldfarb-Shanno method (BFGS). Figure 8.3 presents empirical and fitted variograms of both horizontal components, for 01.03.10–31.05.10.

8.2.2.3 Comparison

In Figure 8.4 we can see that the regression-based method provides a slightly better fit than the generalised method of moments, especially to the X component. The parameters obtained using the regression-based method behave better also out of sample, as Figure 8.5 shows. Table 8.1 compares parameters fitted using these two methods. We notice that while the estimates of σ_X, σ_Y, α_X and α_Y do not differ significantly between the methods, β_X has a different sign depending on the chosen method.

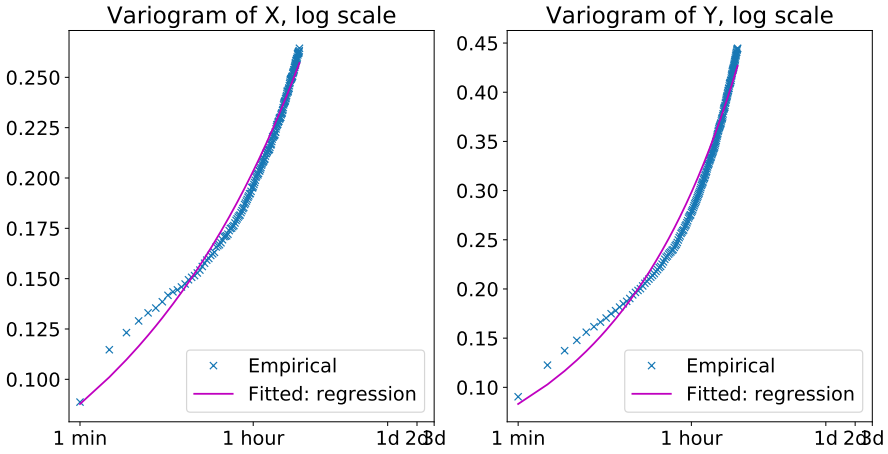


Figure 8.3: Variograms of horizontal components fitted with the regression-based method, using first 180 lags. Training set: 01.03.10–31.05.10.

Method	$\hat{\sigma_X}$	$\hat{\alpha_X}$	$\hat{\beta_X}$	$\hat{\sigma_Y}$	$\hat{\alpha_Y}$	$\hat{\beta_Y}$
GMM	0.114	-0.351	-0.469	0.123	-0.388	0.658
regression	0.107	-0.398	0.072	0.135	-0.345	0.087

Table 8.1: Parameters of horizontal components fitted with the regression-based method and generalised method of moments, using first 180 lags. Training set: 01.03.10–31.05.10.

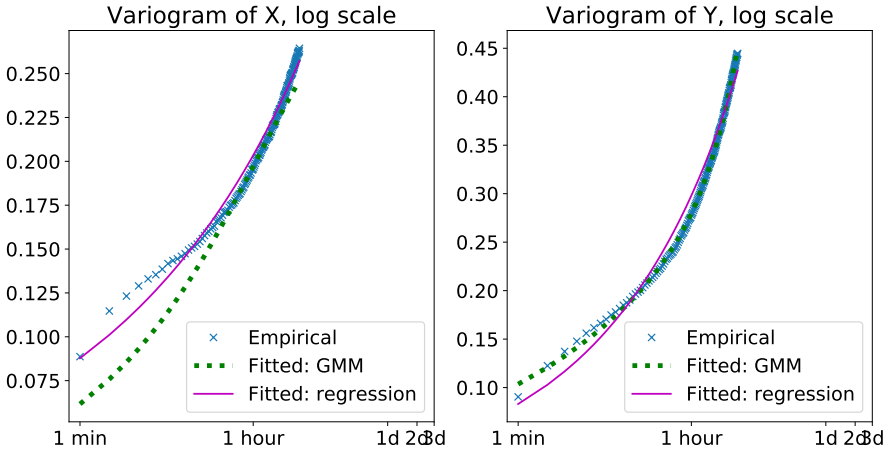


Figure 8.4: Variograms of horizontal components fitted with the regression-based method and generalised method of moments, using first 180 lags. Training set: 01.03.10–31.05.10.

8.3 Forecasting

In this section we propose methods to forecast a single component of wind speed as well as the wind speed itself given past information on both components. In addition to point forecasts, we consider the "ideal" forecasts, which assume the knowledge of the full continuous history, and the "realistic" forecasts based on discrete observations.

8.3.1 Methods

Let $\mathcal{F}$ denote the completed natural filtration of the Brownian motion W. The conditional distribution of a Gaussian moving-average process $Z(t+h)$, defined by Eq. (8.1), given $\mathcal{F}_t$, is Gaussian with mean $\tilde{m}(t,h)$ and variance $\tilde{v}(h)$:

$$\tilde{m}(t,h) = \int_{-\infty}^{t} g(t+h-s)dW(s)$$
$$\tilde{v}(h) = \mathbb{E}\left[\left(\mathbb{E}[Z(t+h)|\mathcal{F}_t] - Z(t+h)\right)^2 |\mathcal{F}_t\right] = \int_0^h g(s)^2 ds.$$

Therefore, the "ideal" point forecast of the moving average process with time horizon h is given by $\tilde{m}(t,h)$ and the L^2 error of this forecast is $\tilde{v}(h)$. We can compare this L^2 error with the L^2 error of the persistence forecast, which assumes that the best predictor of $Z(t+h)$ is $Z(t)$. This error is equal to the variogram

$$v(h) = \mathbb{E}[(Z(t+h) - Z(t))^2] = 2(c(0) - c(h)), \quad h \in \mathbb{R}.$$

We recall that

$$c(t-u) = \int_{-\infty}^{\min\{t,u\}} g(t-s)g(u-s)ds.$$

In practice we aim to forecast $Z(t_n+h)$ from a sequence of discrete observations $Z(t_1), \ldots, Z(t_n)$ using the Gaussian structure of the process (details in [Davis,

2006]). The conditional mean is given by

$$\mathbb{E}[Z(t_n + h)|Z(t_1), \ldots, Z(t_n)] = \mathbf{v}^\top \Omega^{-1} \mathbf{z},$$

where

- $\mathbf{v} \in \mathbb{R}^n$ with $v_i = c(t_n + h - t_i)$;
- $\Omega \in M^n(\mathbb{R})$ with $\Omega_{ij} = c(t_i - t_j)$;
- $\mathbf{z} \in \mathbb{R}^n$ with $z_i = Z(t_i)$.

We can express the conditional variance (and the L^2 error of the point forecast) as

$$\begin{aligned}&\mathbb{E}\left[\left(\mathbb{E}[Z(t_n + h)|Z(t_1), \ldots, Z(t_n)] - Z(t_n + h)\right)^2 |Z(t_1), \ldots, Z(t_n)\right] \\ &= c(0) - \mathbf{v}^\top \Omega^{-1} \mathbf{v}.\end{aligned}$$

When reporting empirical results, in practice we proceed as follows. Taking one minute as our unit, we set:

- $\Delta = |t_i - t_{i+1}|$: the interval between consecutive time steps (in our case $\Delta = 1$).
- n_{dates}: the number of data points used to produce a single forecast (in our case $n_{\text{dates}} = 720$).
- h: the forecast horizon (in our case $h \in \{1, 5, 10, 20, 30, 60, 120, 180, 360\}$ minutes).

We split the data into four three-month-long training and one-month-long testing sets, as specified in Table 8.2. We deseasonalise the testing set with the 2880 dummy variables obtained using the training set, as described in Section 7.2. Therefore we forecast deseasonalised values, which we can easily convert into forecasts with seasonality effects by reverting the deseasonalising and smoothing procedure from Section 7.2.
As before, we approximate all integrals with the function `quad` from Python's

library `scipy.integrate` ([Virtanen et al., 2019]), based on 15-point Gauss-Kronrod quadrature. We compute the theoretical discrete and continuous errors, both for the forecast based on the Gaussian moving average and the persistence forecast. Finally, we report also the actual root mean square errors for both models, i.e. the square root of the sum of squared differences between the actually predicted values and the test data.

Fitting		Forecasting	
Start	End	Start	End
01.12.2009	28.02.2010	01.03.2010	31.03.2010
01.03.2010	31.05.2010	01.06.2010	30.06.2010
01.06.2010	31.08.2010	01.09.2010	30.09.2010
01.09.2010	30.11.2010	01.12.2010	31.12.2010

Table 8.2: Training and testing data sets.

The explicit formulae for conditional means and errors allow us to construct probabilistic forecasts in addition to point forecasts. For each future observation we can construct a prediction interval, i.e. an interval in which this observation is expected to fall with a given probability, given the previous observations. We can express lower and upper bounds of these intervals as

$$\mu(h) \pm z_{\frac{\alpha}{2}}\sigma(h),$$

where $\mu(h)$ and $\sigma(h)$ denote the expected value and the square root of the L^2 error of the point forecast (discrete or continuous), respectively, while $z_{\frac{\alpha}{2}}$ the quantile of the standard normal distribution.

8.3.2 Numerical results

In Figure 8.5 we compare the actual root mean square errors of forecasts based on parameters estimated with the generalised method of moments (Subsection 8.2.1.1) and the regression-based method (Subsection 8.2.1.2) as well as of the persistence forecast. The regression-based method provides better forecasts than the generalised method of moments. The Gaussian

moving average model outperforms the persistence forecast for short time horizons.

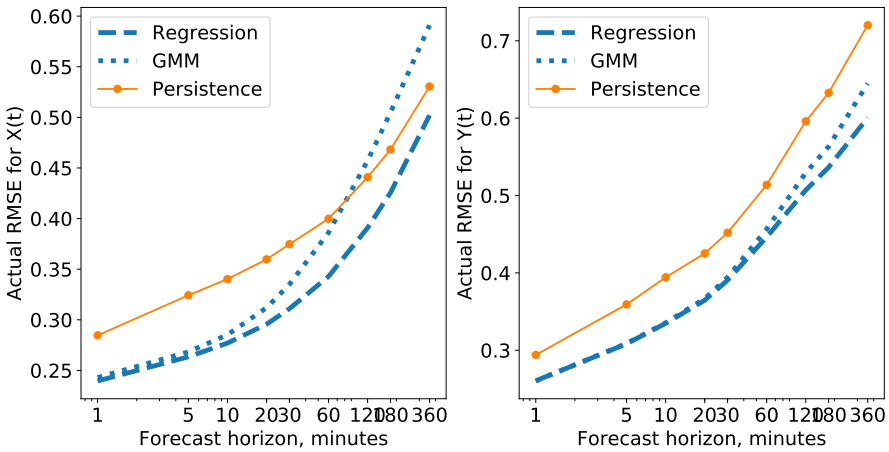


Figure 8.5: Actual root mean square errors of forecasted horizontal components X and Y, compared to persistence forecasts. Parameters estimated with the generalised method of moments (Subsection 8.2.1.1) and the regression-based method (Subsection 8.2.1.2). Training set: 01.03.10–31.05.10. Testing set: 01.06.10–30.06.10.

8.3.2.1 The comparison with gamma and power-law kernels

In Section 8.1 we mentioned other examples of kernels used to model Gaussian moving average processes. One can wonder how the performance of various kernels differs. In this section we compare the models resulting from the triple-scale kernel with gamma and power-law kernels. Precisely, we use the procedure described in Subsection 8.2.1.2 to estimate moving average Gaussian process with the following kernels:

- triple-scale kernel: $g(t) = \sigma t^{\alpha} \left(1 + \frac{t}{60}\right)^{\beta} \exp\left(-\frac{t}{1440}\right)$ for $\alpha > -\frac{1}{2}$, $\beta \in \mathbb{R}$ and $\sigma > 0$;
- gamma kernel: $g_1(t) = \sigma t^{\alpha} \exp\left(-\lambda t\right)$ for $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$ and $\lambda > 0$;
- power-law kernel: $g_2(t) = \sigma t^{\alpha} \left(1 + \frac{t}{60}\right)^{\beta - \alpha}$ for $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$ and $\beta < -\frac{1}{2}$,

where $t > 0$. We note a slightly different parametrisation of the triple-scale and power-law kernels: we keep the parametrisation of the power-law kernel suggested by [Bennedsen et al., 2017].

We start by estimating the parameters of both horizontal components modelled with gamma and power-law kernels based on 180 lags, as we did with the triple-scale kernel. Since in the regression-based estimation procedure σ and α do not depend on the rest of the kernel, their values are the same for all three kernels. To estimate the remaining parameters we use the function `minimize` from Python's library `scipy.optimize` ([Virtanen et al., 2019]) with the limited-memory Broyden-Fletcher-Goldfarb-Shanno method (L-BFGS-B). It results in $\lambda = 0.0$ for gamma kernel and $\beta = -0.5$ for power-law kernel, which correspond to the bounds of both parameters. After studying the behaviour of the minimised function we conclude that the algorithm attempts to reduce both kernels to σt^{α}. We also compare the models out of sample. Figure 8.6 shows that in this case gamma and triple-scale kernels are almost indistinguishable and they both outperform power-law kernel and persistence forecasts.

However, the comparison changes drastically when we look at longer time scales. In Figure 8.7 we show the results of fitting the same models to $10,000$ lags instead of just 180. In this case only the triple-scale kernel can capture various types of variogram behaviour. This shows that the triple-scale kernel truly generalises more standard kernels: it can reduce to simpler gamma or power-law kernels for shorter time scales but is able to capture more complex behaviour when needed.

8.4 Summary and outlook

In this chapter we proposed a one-dimensional model of both horizontal components of the wind speed: a Gaussian moving average process with a novel triple-scale kernel. This kernel replicates the complex behaviour of the empirical variogram function better than more standard kernels such as gamma or power-law. Additionally, it reduces to simpler forms when the variogram's behaviour is less complicated, for example for shorter time scales.

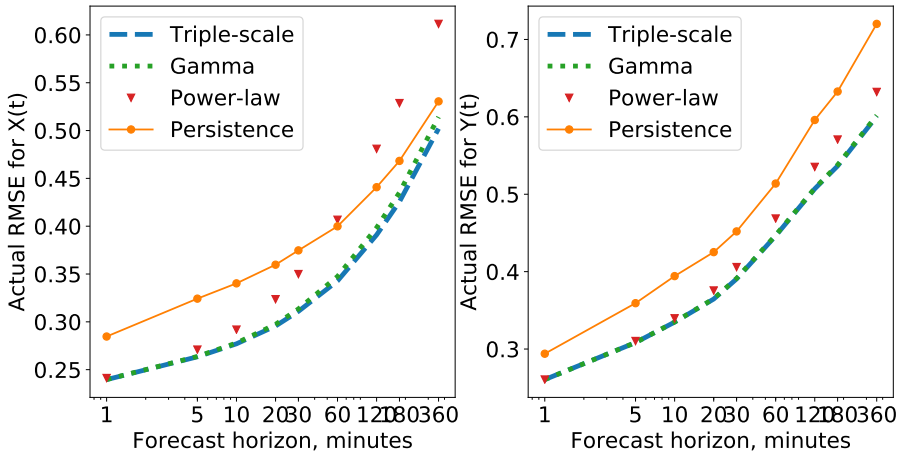


Figure 8.6: The comparison of actual root mean square errors of forecasted horizontal components X and Y with triple-scale, gamma and power-law kernels, and persistence forecasts. Parameters estimated with the regression-based method (Subsection 8.2.1.2). Training set: 01.03.10–31.05.10. Testing set: 01.06.10–30.06.10.

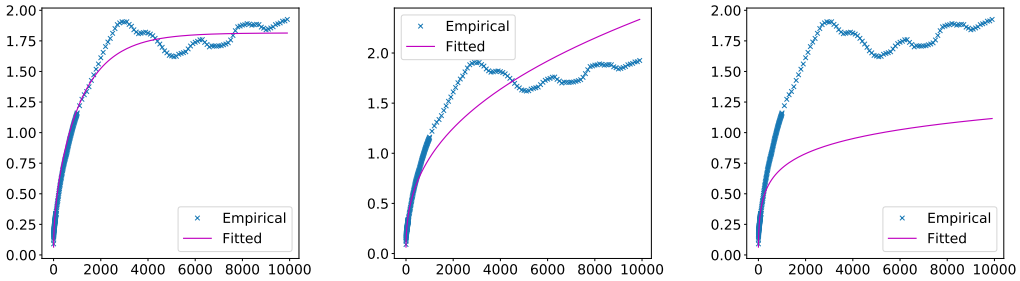


Figure 8.7: Variograms of $Y(t)$ for 01.03.2010–31.05.2010 fitted with the regression-based method (Subsection 8.2.1.2), for lags 1–10,000 (1 lag corresponds to 1 minute). From left to right: triple-scale kernel, gamma kernel and power-law kernel.

We estimated the model using two methods: the generalised method of moments and the regression-based method. We established that the latter provides more accurate estimates. Finally, we produced short-term forecasts of both horizontal components. For both components they outperform the benchmark persistence forecast.

We established that the Brownian semistationary process with the triple-scale kernel is a good model for horizontal components of wind speed. In the next chapter we will explore ways of combining one-dimensional models into

a bivariate model which will allow us to produce short-term forecasts of wind speed.

9

Two-dimensional setting

In Chapter 8 we modelled the horizontal components of wind speeds with Brownian semistationary processes with a novel triple-scale kernel. In this chapter we suggest a way of joining these univariate models into a bivariate wind speed model with correlated marginals. We also develop two approximations to wind speed forecasts and test them against two benchmark models: persistence and direct wind speed prediction.

9.1 Bivariate Gaussian moving average processes

To jointly model and forecast two horizontal components of the wind speed, we must extend the setting of Gaussian moving average process analysed in Section 8.1 to two dimensions. To this end, we propose the following flexible framework. For $t \in \mathbb{R}$ we define a general bivariate Brownian semistationary process by

$$Z(t) = \begin{pmatrix} Z_1(t) \\ Z_2(t) \end{pmatrix} = \int_{-\infty}^{t} K(t-s) d \begin{pmatrix} W_1(s) \\ W_2(s) \end{pmatrix}, \tag{9.1}$$

where for $s \leq t$

$$K(t-s) = \begin{pmatrix} k_{11}(t-s) & k_{12}(t-s) \\ k_{21}(t-s) & k_{22}(t-s) \end{pmatrix}.$$

The Brownian motions W_1 and W_2 are independent, since we include all the dependence structure in the matrix K. Eq. (9.1) allows us to compute the auto- and cross-covariances explicitly.

Proposition 9.0.1. *For $h \geq 0$ and $Z(t)$ as in Eq. (9.1) the following hold.*

1. *The autocovariances of Z_1 and Z_2 have the form:*

$$\begin{aligned} &\mathbb{E}[Z_1(t)Z_1(t+h)] \\ &= \int_{-\infty}^{t} (k_{11}(t-s)k_{11}(t+h-s) + k_{21}(t-s)k_{21}(t+h-s))\, ds, \\ &\mathbb{E}[Z_2(t)Z_2(t+h)] \\ &= \int_{-\infty}^{t} (k_{12}(t-s)k_{12}(t+h-s) + k_{22}(t-s)k_{22}(t+h-s))\, ds. \end{aligned} \tag{9.2}$$

2. *The cross-covariances of Z_1 and Z_2 have the form:*

$$\begin{aligned} &\mathbb{E}[Z_1(t)Z_2(t+h)] \\ &= \int_{-\infty}^{t} (k_{11}(t-s)k_{21}(t+h-s) + k_{12}(t-s)k_{22}(t+h-s))\, ds, \\ &\mathbb{E}[Z_1(t+h)Z_2(t)] \\ &= \int_{-\infty}^{t} (k_{21}(t-s)k_{11}(t+h-s) + k_{22}(t-s)k_{12}(t+h-s))\, ds. \end{aligned} \tag{9.3}$$

Proof. Let us define

$$\begin{aligned} U(t) &= \int_{-\infty}^{t} A(t-s)dW_1(s) + \int_{-\infty}^{t} B(t-s)dW_2(s), \\ V(t) &= \int_{-\infty}^{t} C(t-s)dW_1(s) + \int_{-\infty}^{t} D(t-s)dW_2(s), \end{aligned} \tag{9.4}$$

where W_1, W_2 denote independent standard Brownian motions and A, B, C, D square integrable kernel functions. For $h \geq 0$ we compute

$$\begin{aligned}
&C\left\{(\alpha,\beta)^T;(U(t),V(t+h))^T\right\} = \log\mathbb{E}\left[e^{i(\alpha U(t)+\beta V(t+h))}\right] \\
&= \log\mathbb{E}\left[e^{i\left(\alpha\int_{-\infty}^{t}A(t-s)dW_1(s)+\beta\int_{-\infty}^{t+h}C(t+h-s)dW_1(s)\right)}\right] \\
&+ \log\mathbb{E}\left[e^{i\left(\alpha\int_{-\infty}^{t}B(t-s)dW_2(s)+\beta\int_{-\infty}^{t+h}D(t+h-s)dW_2(s)\right)}\right] =: P+Q,
\end{aligned}$$

where in the first equality we use the definition of cumulant generating function ([Sato et al., 1999]), while in the second one the independence of W_1 and W_2. Using the Lévy–Khintchine formula for Gaussian processes, we express the first factor as

$$\begin{aligned}
P &= \log\mathbb{E}\left[e^{i\int_{-\infty}^{t}\alpha A(t-s)+\beta C(t+h-s)dW_1(s)}\right] + \log\mathbb{E}\left[e^{i\int_{t}^{t+h}\beta C(t+h-s)dW_1(s)}\right] \\
&= \int_{-\infty}^{t} e^{-\frac{1}{2}(\alpha A(t-s)+\beta C(t+h-s))^2}ds + \int_{t}^{t+h} e^{-\frac{1}{2}\beta^2C^2(t+h-s)}ds \\
&= \int_{-\infty}^{t} e^{-\frac{1}{2}\alpha^2A^2(t-s)-\alpha A(t-s)C(t+h-s)}ds + \int_{-\infty}^{t+h} e^{-\frac{1}{2}\beta^2C^2(t+h-s)}ds.
\end{aligned}$$

Now we can differentiate P to obtain

$$\begin{aligned}
&-\left.\frac{\partial^2 P}{\partial\alpha\partial\beta}\right|_{\alpha=0,\beta=0} \\
&= -\frac{\partial}{\partial\beta}\left(\int_{-\infty}^{t} e^{-\frac{1}{2}(\alpha A(t-s)+\beta C(t+h-s))^2}\right. \\
&\left.\left(-\alpha A^2(t-s)-\beta A(t-s)C(t+h-s)\right)ds\right)\Bigg|_{\alpha=0,\beta=0} \\
&= \int_{-\infty}^{t} e^{-\frac{1}{2}(\alpha A(t-s)+\beta C(t+h-s))^2}\left(-\alpha A(t-s)C(t+h-s)\right) \\
&\times\left(-\alpha A^2(t-s)-\beta A(t-s)C(t+h-s)\right)ds\Big|_{\alpha=0,\beta=0} \\
&+ \int_{-\infty}^{t} e^{-\frac{1}{2}(\alpha A(t-s)+\beta C(t+h-s))^2}A(t-s)C(t+h-s)ds\Big|_{\alpha=0,\beta=0} \\
&= \int_{-\infty}^{t} A(t-s)C(t+h-s)ds.
\end{aligned}$$

Similar calculations show that

$$-\frac{\partial^2 Q}{\partial\alpha\partial\beta}\bigg|_{\alpha=0,\beta=0} = \int_{-\infty}^{t} B(t-s)D(t+h-s)ds.$$

Therefore

$$\begin{aligned}
&-\frac{\partial^2}{\partial\alpha\partial\beta}\bigg|_{\alpha=0,\beta=0} C\left\{(\alpha,\beta)^T;(U(t),V(t+h))^T\right\} \\
&= \int_{-\infty}^{t} A(t-s)C(t+h-s) + B(t-s)D(t+h-s)ds.
\end{aligned}$$

Finally, we substitute the terms in Eq. (9.4).

1. $A(t-s) = k_{11}(t-s)$, $B(t-s) = k_{12}(t-s)$, $C(t-s) = k_{21}(t-s)$ and $D(t-s) = k_{22}(t-s)$:

$$\begin{aligned}
\mathbb{E}[Z_1(t)Z_2(t+h)] &= -\frac{\partial^2}{\partial\alpha\partial\beta}\bigg|_{\alpha=0,\beta=0} C\left\{(\alpha,\beta)^T;(Z_1(t),Z_2(t+h))^T\right\} \\
&= \int_{-\infty}^{t} k_{11}(t-s)k_{21}(t+h-s) + k_{12}(t-s)k_{22}(t+h-s)ds.
\end{aligned}$$

2. $A(t-s) = k_{21}(t-s)$, $B(t-s) = k_{22}(t-s)$, $C(t-s) = k_{11}(t-s)$ and $D(t-s) = k_{12}(t-s)$:

$$\begin{aligned}
\mathbb{E}[Z_1(t+h)Z_2(t)] &= -\frac{\partial^2}{\partial\alpha\partial\beta}\bigg|_{\alpha=0,\beta=0} C\left\{(\alpha,\beta)^T;(Z_2(t),Z_1(t+h))^T\right\} \\
&= \int_{-\infty}^{t} k_{21}(t-s)k_{11}(t+h-s) + k_{22}(t-s)k_{12}(t+h-s)ds.
\end{aligned}$$

3. $A(t-s) = k_{11}(t-s) = C(t-s)$, $B(t-s) = k_{12}(t-s) = D(t-s)$:

$$\begin{aligned}
\mathbb{E}[Z_1(t)Z_1(t+h)] &= -\frac{\partial^2}{\partial\alpha\partial\beta}\bigg|_{\alpha=0,\beta=0} C\left\{(\alpha,\beta)^T;(Z_1(t),Z_1(t+h))^T\right\} \\
&= \int_{-\infty}^{t} k_{11}(t-s)k_{11}(t+h-s) + k_{12}(t-s)k_{12}(t+h-s)ds.
\end{aligned}$$

4. $A(t-s) = k_{22}(t-s) = C(t-s)$, $B(t-s) = k_{21}(t-s) = D(t-s)$:

$$\begin{aligned}\mathbb{E}[Z_2(t)Z_2(t+h)] &= -\frac{\partial^2}{\partial\alpha\partial\beta}\bigg|_{\alpha=0,\beta=0} C\left\{(\alpha,\beta)^T;(Z_2(t),Z_2(t+h))^T\right\}\\ &= \int_{-\infty}^{t} k_{22}(t-s)k_{22}(t+h-s) + k_{21}(t-s)k_{21}(t+h-s)ds.\end{aligned}$$

□

We consider a two-dimensional model which accounts for the non-zero correlation between horizontal components. Precisely, we use Eq. (9.1) with $W_1 = W^X$, $W_2 = W^Y$ and a kernel

$$K(t-s) = \begin{pmatrix} g_X(t-s) & 0 \\ \rho g_X(t-s) & \sqrt{1-\rho^2}g_Y(t-s) \end{pmatrix},$$

where $t \geq s$, g_X and g_Y denote square integrable functions, W^X and W^Y independent standard Brownian motions, while ρ a constant correlation coefficient. We use the convention that $g_X(t) = g_Y(t) = 0$ for $t \leq 0$ and $g_X(t) \geq 0$, $g_Y(t) \geq 0$ for $t > 0$. Now $X(t)$ and $Y(t)$ become

$$\begin{cases} X(t) = \int_{-\infty}^{t} g_X(t-s)dW^X(s), \\ Y(t) = \int_{-\infty}^{t} g_Y(t-s)dW^Y(s), \\ dW^X(s)\,dW^Y(s) = \rho dt. \end{cases} \tag{9.5}$$

Proposition 9.0.1 gives us the forms of autocovariances and cross-covariances of Eq. (9.5) for $h \geq 0$:

$$\begin{aligned}\mathbb{E}[X(t)X(t+h)] &= \int_0^\infty g_X(s)g_X(s+h)ds,\\ \mathbb{E}[Y(t)Y(t+h)] &= \int_0^\infty g_Y(s)g_Y(s+h)ds,\\ \mathbb{E}[X(t)Y(t+h)] &= \rho\int_0^\infty g_X(s)g_Y(s+h)ds,\\ \mathbb{E}[Y(t)X(t+h)] &= \rho\int_0^\infty g_Y(s)g_X(s+h)ds.\end{aligned} \tag{9.6}$$

The model Eq. (9.5) allows for an asymmetric cross-covariance, i.e. in general for $h \neq 0$

$$\mathbb{E}[X(t)Y(t+h)] \neq \mathbb{E}[X(t+h)Y(t)].$$

We also define a two-dimensional analogue of the variogram: the *cross-variogram*:

$$\begin{aligned}
&\mathbb{E}\left[(X(t+h)-X(t))(Y(t+h)-Y(t))\right] \\
&= \rho\left(2\int_0^\infty g_X(s)g_Y(s)ds - \int_0^\infty g_X(s+h)g_Y(s)ds - \int_0^\infty g_X(s)g_Y(s+h)ds\right).
\end{aligned} \tag{9.7}$$

9.2 Estimation

In this section we suggest a novel calibration method which combines two procedures described in Section 8.2: the generalised method of moments and the regression-based method. Our aim is to capture the behaviour of marginals as well as the correlation structure.

9.2.1 Methods

We estimate seven parameters $(\rho, \sigma_X, \alpha_X, \beta_X, \sigma_Y, \alpha_Y, \beta_Y)$ as follows. We use the regression-based procedure (Subsection 8.2.1.2) to compute $\hat{\sigma_X}$, $\hat{\alpha_X}$, $\hat{\sigma_Y}$ and $\hat{\alpha_Y}$, the estimates of σ_X, α_X, σ_Y and α_Y, respectively. We set $(\rho^0, \beta_X^0, \beta_Y^0) = (0.0, 0.0, 0.0)$ and estimate the remaining parameters (ρ, β_X, β_Y) by matching the empirical and theoretical weighted cross-variograms and variograms. Precisely, we minimise the expression

$$\begin{aligned}
&w\left(\mathbb{E}\left[(X(t+h)-X(t))(Y(t+h)-Y(t))\right]\right. \\
&\quad \left.-\hat{\mathbb{E}}\left[(\hat{X}(t+h)-\hat{X}(t))(\hat{Y}(t+h)-\hat{Y}(t))\right]\right)^2 \\
&+\left(\mathbb{E}\left[(X(t+h)-X(t))^2\right) - \hat{\mathbb{E}}\left[(\hat{X}(t+h)-\hat{X}(t))^2\right)\right)^2 \\
&+\left(\mathbb{E}\left[(Y(t+h)-Y(t))^2\right] - \hat{\mathbb{E}}\left[(\hat{Y}(t+h)-\hat{Y}(t))^2\right]\right)^2,
\end{aligned} \tag{9.8}$$

where $\hat{X}$ and $\hat{Y}$ denote the empirical data and $h \in \{1, 5, 10, 20, 30, 60, 120, 180, 360\}$ (minutes). Let

$$f_1^{\mathrm{GMM}}(X, \theta, h) := (X(t+h) - X(t))^2) - (\hat{X}(t+h) - \hat{X}(t))^2$$

and

$$\begin{aligned} &f_2^{\mathrm{GMM}}(X, Y, \theta, h) \\ &:= (X(t+h) - X(t))(Y(t+h) - Y(t)) - (\hat{X}(t+h) - \hat{X}(t))(\hat{Y}(t+h) - \hat{Y}(t)). \end{aligned}$$

Then our method of parameter estimation corresponds to the one-step generalised method of moments (Subsection 8.2.1.1) with the weight matrix

$$W = \left[\begin{array}{c|c|c} I & 0 & 0 \\ \hline 0 & wI & 0 \\ \hline 0 & 0 & I \end{array}\right]$$

and the function

$$f^{\mathrm{GMM}}(Z, \theta) = \begin{bmatrix} f_1^{\mathrm{GMM}}(X, \theta, 1) \\ \dots \\ f_1^{\mathrm{GMM}}(X, \theta, 360) \\ f_2^{\mathrm{GMM}}(X, Y, \theta, 1) \\ \dots \\ f_2^{\mathrm{GMM}}(X, Y, \theta, 360) \\ f_1^{\mathrm{GMM}}(Y, \theta, 1) \\ \dots \\ f_1^{\mathrm{GMM}}(Y, \theta, 360) \end{bmatrix},$$

where I denotes the identity matrix of dimension 9×9 (9 is the length of the vector of considered forecast horizons h). We start the least squares procedure with $(\rho^0, \beta_X^0, \beta_Y^0) = (0.0, 0.0, 0.0)$. We use the function `minimize` from Python's library `scipy.optimize` ([Virtanen et al., 2019]) with Sequential Least Squares Programming (SLSQP) method and bounds $(\rho, \beta_X, \beta_Y) \in (-1.0, 1.0) \times (-\infty, \infty) \times (-\infty, \infty)$. We run the procedure for different weights w and numbers of lags.

Remark 9.1. Alternatively, we could use the generalised method of moments to estimate all seven parameters. However, the computational time significantly (more than ten times) exceeds the computational time of the previous method. Moreover, the joint estimation of all seven parameters is not computationally stable, even when we set $\hat{\sigma_X}$, $\hat{\alpha_X}$, $\hat{\sigma_Y}$ and $\hat{\alpha_Y}$ estimated using the regression-based method as initial parameters. Therefore we recommend the combination of the regression-based method and the generalised method of moments described above.

9.2.2 Numerical results

In this section we present the results of full model estimation as described in Section 9.2. We use the first 180 lags since we are interested in short-term forecasts, mostly up to 3 hours (180 minutes). We choose different values of the weight w.

9.2.2.1 Weight $w = 1$

Table 9.1 shows the estimates of all seven parameters. The parameters estimated with the regression-based method ($\hat{\sigma_X}$, $\hat{\alpha_X}$, $\hat{\sigma_Y}$ and $\hat{\alpha_Y}$) do not vary much for different time periods. On the other hand, the estimates of the remaining parameters are not consistent among seasons, which most likely indicates problems with parameter identification in the generalised method of moments. The issue could result from some seasonal effects remaining even after the deseasonalisation procedure. Finally, it is possible the suggested model cannot fully capture the behaviour of our data set.
Nevertheless, our estimation method provides good fits to variograms, as presented in Figure 9.1. However, Figure 9.2 suggests that we could improve the fit to the cross-variogram.

	$\hat{\rho}$	$\hat{\sigma_X}$	$\hat{\alpha_X}$	$\hat{\beta_X}$	$\hat{\sigma_Y}$	$\hat{\alpha_Y}$	$\hat{\beta_Y}$
01.12.2009–28.02.2010	0.075	0.107	-0.373	0.146	0.120	-0.317	0.124
01.03.2010–31.05.2010	0.024	0.107	-0.398	0.072	0.135	-0.345	0.087
01.06.2010–31.08.2010	-0.048	0.136	-0.360	0.055	0.173	-0.336	0.030
01.09.2010–30.11.2010	0.020	0.111	-0.373	0.105	0.120	-0.336	0.086

Table 9.1: Fitted parameters for different sets of training data, estimated with the procedure described in Section 9.2, for lags 1–180 and with weight $w = 1$.

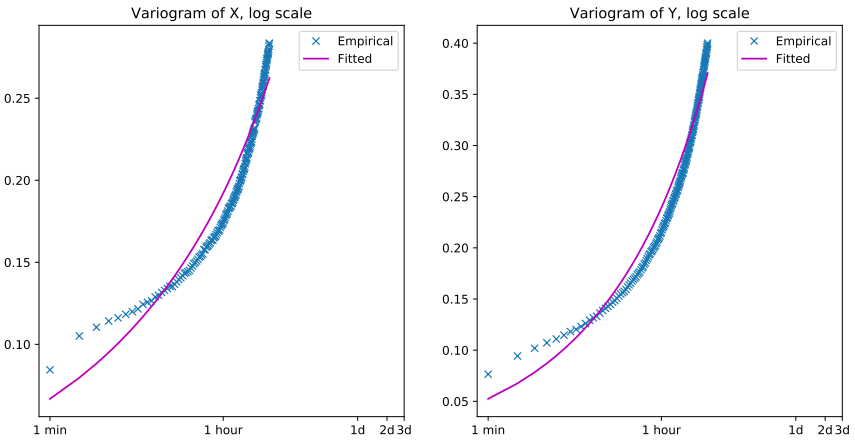


Figure 9.1: Variograms of $X(t)$ and $Y(t)$ for 01.12.2009–28.02.2010, fitted with the procedure described in Section 9.2, for lags 1–180 and with weight $w = 1$. Please note the logarithmic scale of the plots.

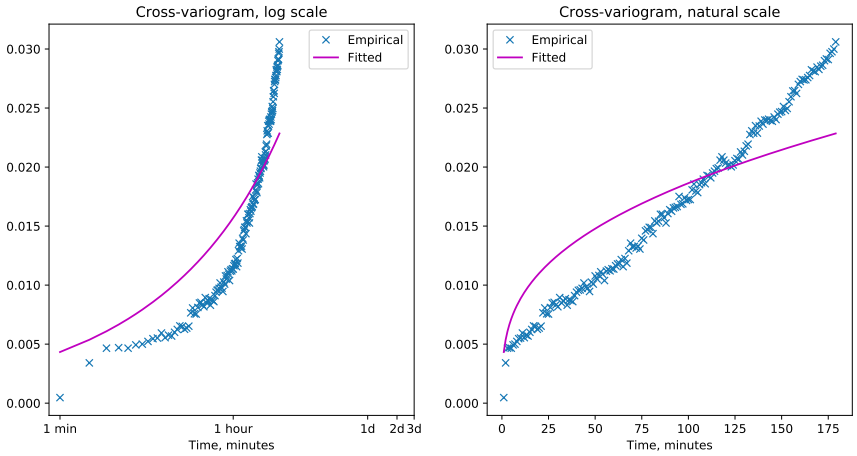


Figure 9.2: The cross-variogram of $X(t)$ and $Y(t)$ for 01.12.2009–28.02.2010,fitted with the procedure described in Section 9.2, for lags 1–180 and with weight $w = 1$. Left: logarithmic scale. Right: natural scale.

9.2.2.2 Weight $w = 100$

Table 9.2, Figure 9.3 and Figure 9.4 indicate that the fit resulting from setting $w = 100$ is similar to $w = 1$, so we suggest setting lower values of w. Throughout the rest of this chapter we will use $w = 1$.

	$\hat{\rho}$	$\hat{\sigma_X}$	$\hat{\alpha_X}$	$\hat{\beta_X}$	$\hat{\sigma_Y}$	$\hat{\alpha_Y}$	$\hat{\beta_Y}$
01.12.2009–28.02.2010	0.073	0.107	-0.373	0.260	0.120	-0.317	0.159
01.03.2010–31.05.2010	0.024	0.107	-0.398	0.097	0.135	-0.345	0.091
01.06.2010–31.08.2010	-0.047	0.136	-0.360	0.127	0.173	-0.336	0.052
01.09.2010–30.11.2010	0.020	0.111	-0.373	0.105	0.120	-0.336	0.086

Table 9.2: Fitted parameters for different sets of training data, estimated with the procedure described in Section 9.2, for lags 1–180 and with weight $w = 100$.

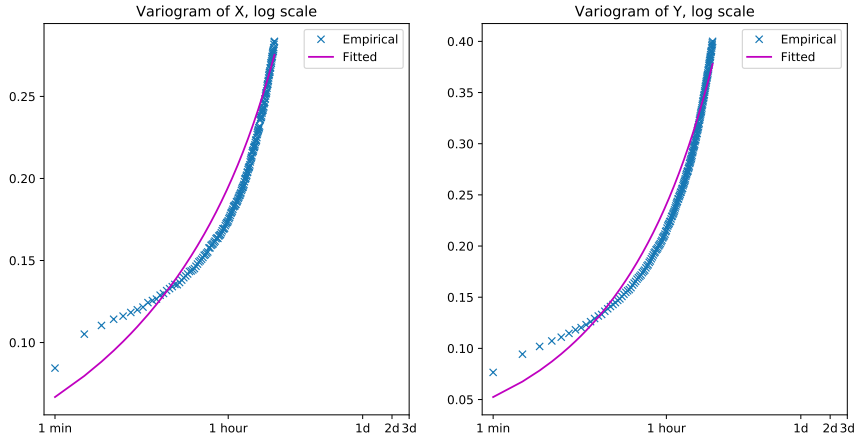


Figure 9.3: Variograms of $X(t)$ and $Y(t)$ for 01.12.2009–28.02.2010, fitted with the procedure described in Section 9.2, for lags 1–180 and with weight $w = 100$. Please note the logarithmic scale of the plots.

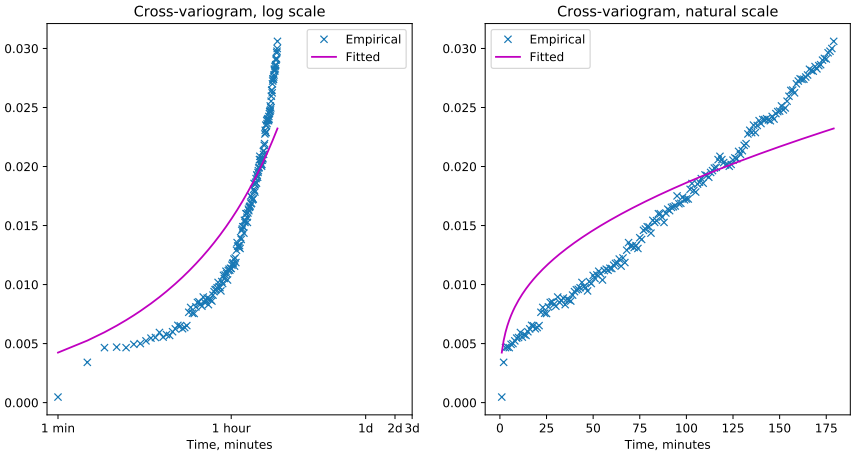


Figure 9.4: The cross-variogram of $X(t)$ and $Y(t)$ for 01.12.2009–28.02.2010, fitted with the procedure described in Section 9.2, for lags 1–180 and with weight $w = 100$. Left: logarithmic scale. Right: natural scale.

9.2.2.3 Maximum likelihood

We explore also the maximum-likelihood method of parameter estimation, very popular in the Gaussian setting (see e.g. [Davis, 2006]). We observe two evenly spaced time series: $\{X_i\}_{i=1}^N$ and $\{Y_i\}_{i=1}^N$, where $X_i := X(i\Delta)$ and $Y_i := Y(i\Delta)$ for $i = 1, 2, \ldots, N$ and $\Delta > 0$. We can treat them as N observations of bivariate pairs $Z_i = (X_i, Y_i)$, $i = 1, \ldots, N$ with the pairwise log-likelihood function

$$l := \sum_{1 \le i < j \le N} f_{(Z_i, Z_j)}(z_i, z_j),$$

where

$$f_{(Z_i,Z_j)}(z_i, z_j) = \frac{1}{2\pi} \left(\det \Sigma_{ij}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \begin{pmatrix} Z_i \\ Z_j \end{pmatrix}^T \Sigma_{ij}^{-\frac{1}{2}} \begin{pmatrix} Z_i \\ Z_j \end{pmatrix}\right)$$

and

$$\begin{aligned} \Sigma_{ij} = \mathrm{Cov}\left(Z_i, Z_j\right) &= \mathrm{Cov}\left(\begin{pmatrix} X_i \\ Y_i \\ X_j \\ Y_j \end{pmatrix}\right) \\ &= \begin{pmatrix} \mathrm{Var}\left(X_i\right) & \mathrm{Cov}\left(X_i, Y_i\right) & \mathrm{Cov}\left(X_i, X_j\right) & \mathrm{Cov}\left(X_i, Y_j\right) \\ \mathrm{Cov}\left(Y_i, X_i\right) & \mathrm{Var}\left(Y_i\right) & \mathrm{Cov}\left(Y_i, X_j\right) & \mathrm{Cov}\left(Y_i, Y_j\right) \\ \mathrm{Cov}\left(X_j, X_i\right) & \mathrm{Cov}\left(X_j, Y_i\right) & \mathrm{Var}\left(X_j\right) & \mathrm{Cov}\left(X_j, Y_j\right) \\ \mathrm{Cov}\left(Y_j, X_i\right) & \mathrm{Cov}\left(Y_j, Y_i\right) & \mathrm{Cov}\left(Y_j, X_j\right) & \mathrm{Var}\left(Y_j\right) \end{pmatrix}. \end{aligned}$$

Since Σ_{ij} is symmetric, we have to compute only the upper or lower triangular part. The explicit formulae for the entries are given by Eq. (9.2). We can estimate the parameters of X and Y by minimising the the log-likelihood l. Unfortunately, in practice the computational cost of this method discourages from its application.

9.3 Forecasting

Because we do not know the explicit formula for $\mathbb{E}[S(t+h)|\mathcal{F}_t]$, we approximate it in two different ways and compare predicted wind speed to test data obtained from deseasonalised horizontal components as $S(t) = \sqrt{X(t)^2 + Y(t)^2}$. Precisely, we first deseasonalise horizontal components in the testing data set with 2880 variables obtained using the procedure described in Section 7.2. Then we use these components to compute deseasonalised wind speed. Therefore both in the estimation and forecasting steps we work on the deseasonalised data.

9.3.1 Methods

In this chapter we focus on point forecasts. We evaluate their forecasting performance by root mean square errors (RMSEs), i.e. the quadratic mean of differences between predicted and actual values at a given time horizon. We compare these errors to RMSEs of simple persistence (sometimes called "no-change") forecasts, since beating the performance of persistence forecasts should constitute the first goal of any forecasting exercise.
[Giebel, 2003] suggest also other loss functions widely used in the forecasting literature. Examples include:

- Root mean square error (RMSE): $\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^{\text{pred}} - y_i^{\text{obs}}\right)^2}$,
- Mean absolute error (MAE): $\frac{1}{N}\sum_{i=1}^{N}\left|y_i^{\text{pred}} - y_i^{\text{obs}}\right|$,
- Mean error (ME): $\frac{1}{N}\sum_{i=1}^{N}\left(y_i^{\text{pred}} - y_i^{\text{obs}}\right)$,
- Histogram of the error distribution,
- Coefficient of determinant (R^2),

where we predict a time series $\{y_i^{\text{pred}}\}_{i=1}^N$ and observe $\{y_i^{\text{obs}}\}_{i=1}^N$.
While in this chapter we focus on forecasting wind speeds, [Hering and Genton,

2010] argue that practitioners would most likely convert these results into predicted wind power production levels using a power curve

$$P = \frac{1}{2}\alpha\rho\pi r^2 S^3,$$

where α denotes a dimensionless efficiency constant of a given turbine type, ρ the air density, r the radius swept by turbine blades and S the wind speed. Loss functions applied to the predicted power rather than wind speeds put more weight on average values, ignoring the extremes. Very low wind speeds do not suffice to start the turbine, while very high force it to stop, therefore within these two extreme regions the actual value of the speed does not matter.

9.3.1.1 Taylor series approximation

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function at the point $\mathbf{a} \in \mathbb{R}^2$. Let $\mathbf{v} = (v_1, v_2)^T$, $\mathbf{x} = (x_1, x_2)^T$, $\mathbf{a} = (a_1, a_2)^T$ and $\mathbf{v} = \mathbf{x} - \mathbf{a}$. Then we can expand

$$\begin{aligned} f(\mathbf{x}) = f(\mathbf{a}) &+ \frac{\partial f}{\partial x_1}(\mathbf{a})v_1 + \frac{\partial f}{\partial x_2}(\mathbf{a})v_2 + \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a})\frac{v_1^2}{2!} + \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a})v_1 v_2 \\ &+ \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a})\frac{v_2^2}{2!} + R_3(\mathbf{x}, \mathbf{a}), \end{aligned}$$

where

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{R_3(\mathbf{x}, \mathbf{a})}{(\mathbf{x} - \mathbf{a})^3} = 0.$$

Let us consider an example which will help us approximate two-dimensional wind speed forecasts.

Example 9.1. Let $\mathbf{x} \in \mathbb{R}^2$, $\mathbf{a} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$, $\mathbf{v} = \mathbf{x} - \mathbf{a}$ and $f(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$. Then

$$\begin{aligned}
\frac{\partial f}{\partial x_1}(\mathbf{a}) &= \frac{a_1}{(a_1^2 + a_2^2)^{\frac{1}{2}}} \\
\frac{\partial f}{\partial x_2}(\mathbf{a}) &= \frac{a_2}{(a_1^2 + a_2^2)^{\frac{1}{2}}} \\
\frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) &= \frac{a_2^2}{2\,(a_1^2 + a_2^2)^{\frac{3}{2}}} \\
\frac{\partial^2 f}{\partial x_1 x_2}(\mathbf{a}) &= -\frac{a_1 a_2}{(a_1^2 + a_2^2)^{\frac{3}{2}}} \\
\frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) &= \frac{a_1^2}{2\,(a_1^2 + a_2^2)^{\frac{3}{2}}}.
\end{aligned}$$

Therefore

$$f(\mathbf{x}) = \left(a_1^2 + a_2^2\right)^{\frac{1}{2}} + \frac{a_1 v_1 + a_2 v_2}{(a_1^2 + a_2^2)^{\frac{1}{2}}} + \frac{a_2^2 v_1^2 - 2a_1 a_2 v_1 v_2 + a_1^2 v_2^2}{2\,(a_1^2 + a_2^2)^{\frac{3}{2}}} + R_3(\mathbf{x}, \mathbf{a}),$$

where

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{R_3(\mathbf{x}, \mathbf{a})}{(\mathbf{x} - \mathbf{a})^3} = 0.$$

Now let $X(t_1), \ldots, X(t_n)$ and $Y(t_1), \ldots, Y(t_n)$ be sequences of discrete observations of our Gaussian processes. We define

$$\begin{aligned}
Z_1 &:= (X(t_n + h), Y(t_n + h))^T, \\
Z_2 &:= (X(t_1), \ldots, X(t_n), Y(t_1), \ldots, Y(t_n))^T.
\end{aligned}$$

Then

$$\mathbb{E}\left[Z_1 | Z_2\right] = \Sigma_{12}\Sigma_{22}^{-1} Z_2 := \begin{bmatrix} \mu_X(h) \\ \mu_Y(h) \end{bmatrix},$$

$$\operatorname{Cov}\left(Z_1 | Z_2\right) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} := \begin{bmatrix} \nu_X(h) & \nu_{XY}(h) \\ \nu_{XY}(h) & \nu_Y(h) \end{bmatrix},$$

where

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \operatorname{Cov}(Z_1) & \operatorname{Cov}(Z_1, Z_2) \\ \operatorname{Cov}(Z_2, Z_1) & \operatorname{Cov}(Z_2) \end{bmatrix}.$$

We can use Example 9.1 to approximate

$$\begin{aligned} S &= \left(X^2+Y^2\right)^{\frac{1}{2}} \\ &\approx \frac{X_0\left(X-X_0\right)+Y_0\left(Y-Y_0\right)}{\left(X_0^2+Y_0^2\right)^{\frac{1}{2}}} \\ &+ \frac{Y_0^2\left(X-X_0\right)^2-2X_0Y_0\left(X-X_0\right)\left(Y-Y_0\right)+X_0^2\left(Y-Y_0\right)^2}{2\left(X_0^2+Y_0^2\right)^{\frac{3}{2}}}. \end{aligned} \tag{9.9}$$

After applying Eq. (9.9) with $(X_0, Y_0) = (\mu_X(h), \mu_Y(h))$ we obtain

$$\begin{aligned} S = \left(X^2+Y^2\right)^{\frac{1}{2}} &\approx \left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{1}{2}} \\ &+ \left(\frac{\mu_X(h)}{\left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{1}{2}}}\left(X-\mu_X(h)\right)+\frac{\mu_Y(h)}{\left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{1}{2}}}\left(Y-\mu_Y(h)\right)\right) \\ &+ \frac{1}{2}\left(\frac{\mu_Y(h)^2}{\left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{3}{2}}}\left(X-\mu_X(h)\right)^2\right. \\ &- \frac{2\mu_X(h)\mu_Y(h)}{\left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{3}{2}}}\left(X-\mu_X(h)\right)\left(Y-\mu_Y(h)\right) \\ &+ \left.\frac{\mu_X(h)^2}{\left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{3}{2}}}\left(Y-\mu_Y(h)\right)^2\right). \end{aligned}$$

Now we can take conditional expectations of both sides to get

$$\begin{aligned} \mathbb{E}\left[S(t_n+h)|Z_2\right] &= \mathbb{E}\left[\left(X(t_n+h)^2+Y(t_n+h)^2\right)^{\frac{1}{2}}|Z_2)\right] \\ &\approx \left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{1}{2}} + \frac{1}{2}\left(\frac{\mu_Y(h)^2}{\left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{3}{2}}}\nu_X(h)\right. \\ &- \left.\frac{2\mu_X(h)\mu_Y(h)}{\left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{3}{2}}}\nu_{XY}(h) + \frac{\mu_X(h)^2}{\left(\mu_X(h)^2+\mu_Y(h)^2\right)^{\frac{3}{2}}}\nu_Y(h)\right). \end{aligned}$$

9.3.1.2 Monte Carlo approximation

Let us define

$$\begin{bmatrix} X(t_n+h|t_1,\ldots,t_n) \\ Y(t_n+h|t_1,\ldots,t_n) \end{bmatrix} := \begin{bmatrix} X(t_n+h) \\ Y(t_n+h) \end{bmatrix} \Big| \begin{bmatrix} X(t_1) \\ Y(t_1) \end{bmatrix}, \ldots, \begin{bmatrix} X(t_n) \\ Y(t_n) \end{bmatrix}$$

With notation the same as in the Taylor series approximation example, we know that

$$\begin{bmatrix} X(t_n+h|t_1,\ldots,t_n) \\ Y(t_n+h|t_1,\ldots,t_n) \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_X(h) \\ \mu_Y(h) \end{bmatrix}, \begin{bmatrix} \nu_X(h) & \nu_{XY}(h) \\ \nu_{XY}(h) & \nu_Y(h) \end{bmatrix} \right). \tag{9.10}$$

Therefore we can simulate $N_{\text{sim}} = 10,000$ variables

$$\begin{bmatrix} X(t_n+h|t_1,\ldots,t_n) \\ Y(t_n+h|t_1,\ldots,t_n) \end{bmatrix}_{\text{sim}}^{1}, \ldots, \begin{bmatrix} X(t_n+h|t_1,\ldots,t_n) \\ Y(t_n+h|t_1,\ldots,t_n) \end{bmatrix}_{\text{sim}}^{N_{\text{sim}}}$$

from the bivariate normal distribution given by Eq. (9.10) and approximate

$$\begin{aligned} &\mathbb{E}\left[S(t_n+h)|X(t_1),\ldots,X(t_n),Y(t_1),\ldots,Y(t_n)\right] \\ &\approx \frac{1}{N_{\text{sim}}} \Sigma_{i=1}^{N_{\text{sim}}} \sqrt{\left(X_{\text{sim}}^i(t_n+h|t_1,\ldots,t_n)\right)^2 + \left(Y_{\text{sim}}^i(t_n+h|t_1,\ldots,t_n)\right)^2}. \end{aligned} \tag{9.11}$$

9.3.2 Numerical results

We compare the persistence forecast with two approximations of forecasts based on the Gaussian moving average model. Table 8.2 presents three-month-long training and one-month-long testing data sets. The models are calibrated on training sets with the method described in Section 9.2, with $w = 1$ and first 180 lags. The maximum forecasting horizon is 360 minutes. Figure 9.5 presents root mean square errors of wind speed forecasts based on the Gaussian moving average model, approximated via Taylor series and $N_{\text{sim}} = 10,000$ Monte Carlo simulations, compared to the persistence forecast. We described the method of computation of these errors in Section 8.3. Both approximations of speed forecasts give very similar results. Up to about three

hours the Gaussian moving average model performs significantly better than the persistence forecast. As we can see in Figure 9.6 and Figure 9.7, the Gaussian moving average model outperforms the persistence forecast also for horizontal components.

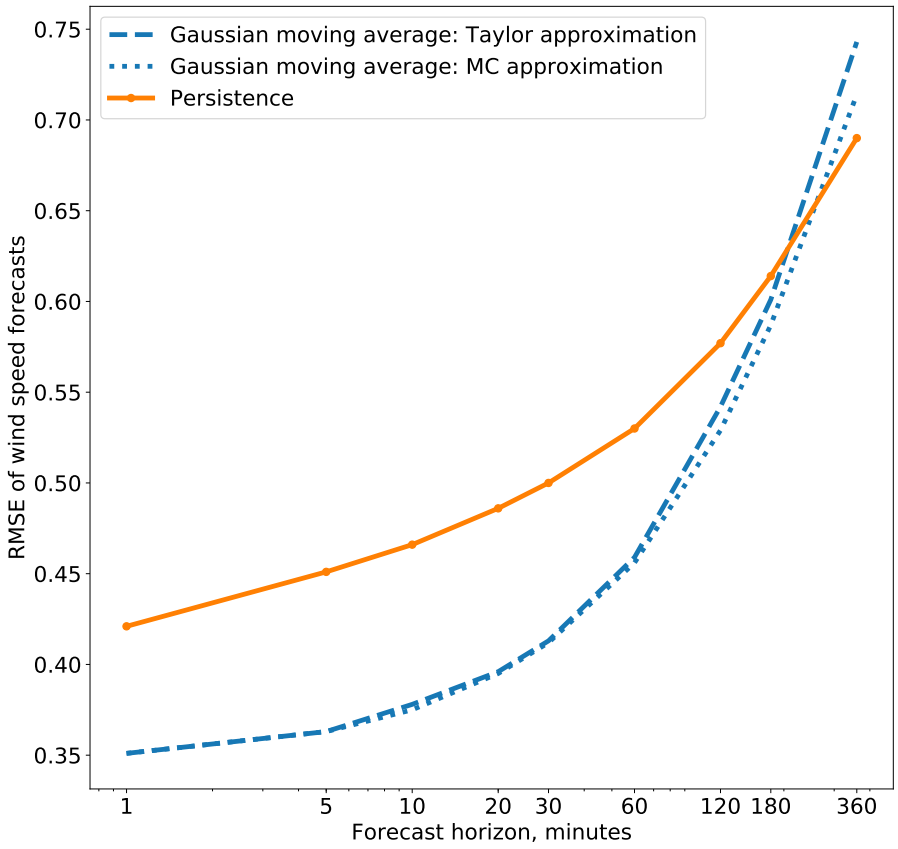


Figure 9.5: Root mean square errors of forecasted wind speed S, approximated with a Taylor series and Monte Carlo simulation, compared to the persistence forecast. Models calibrated with the method described in Section 9.2, with $w = 1$ and first 180 lags. Training set: 01.12.09–28.02.10, testing set: 01.03.10–31.03.10.

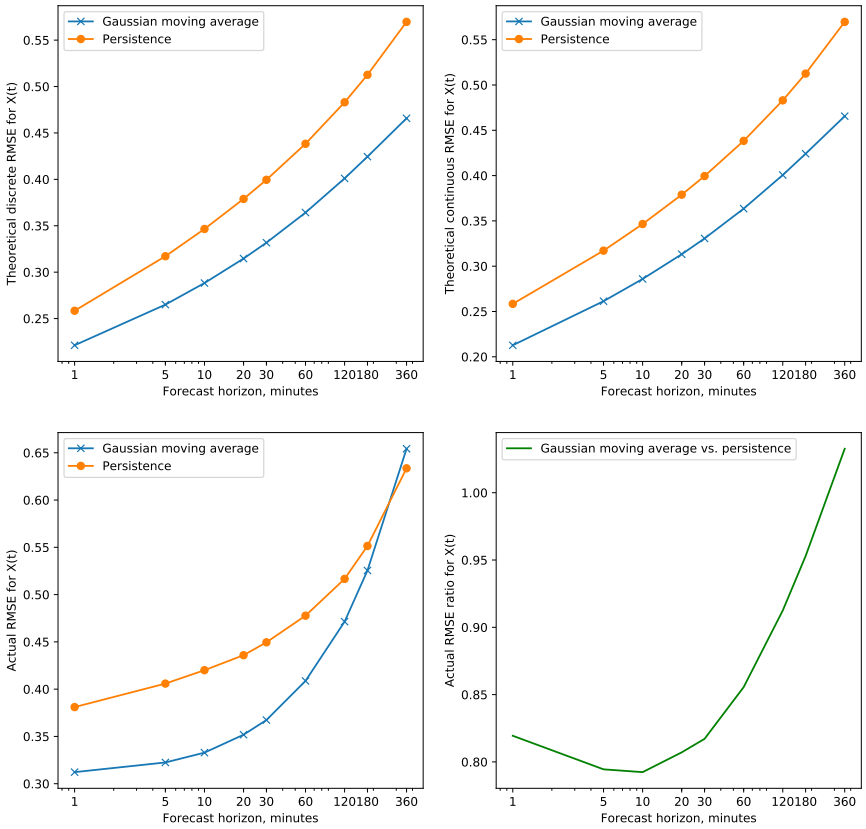


Figure 9.6: Root mean square errors of forecasted horizontal component X: theoretical continuous, theoretical discrete and actual, compared to persistence forecasts. The bottom right plot: the ratio between the actual Gaussian moving average forecast and persistence forecast. Models calibrated with the method described in Section 9.2, with $w = 1$ and first 180 lags. Training set: 01.12.09–28.02.10, testing set: 01.03.10–31.03.10.

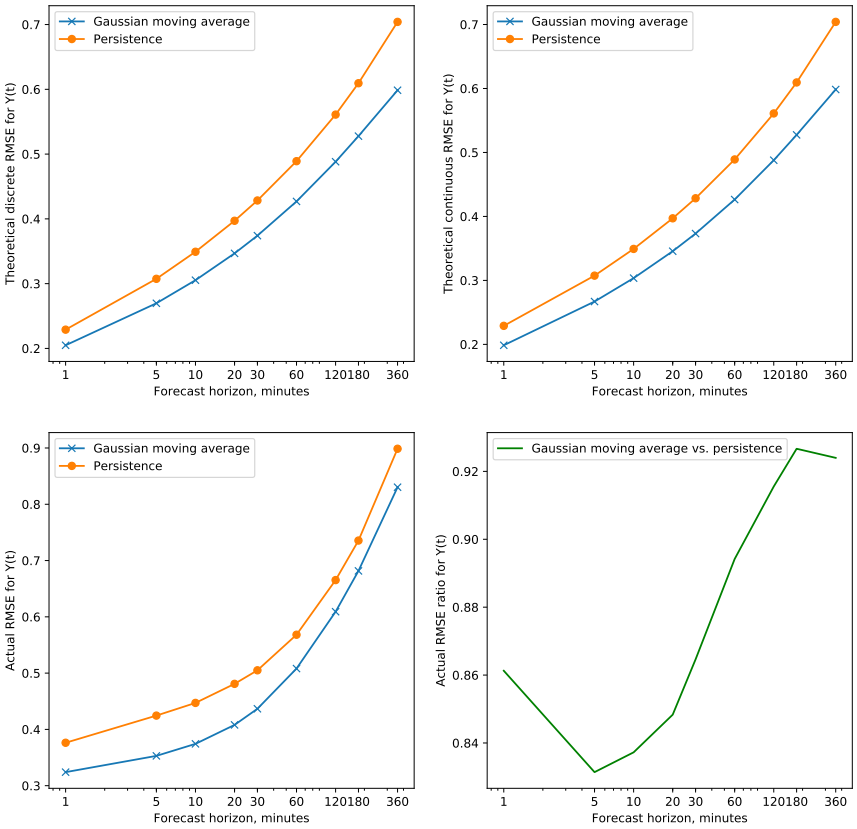


Figure 9.7: Root mean square errors of forecasted horizontal component Y: theoretical continuous, theoretical discrete and actual, compared to persistence forecasts. The bottom right plot: the ratio between the actual Gaussian moving average forecast and persistence forecast. Models calibrated with the method described in Section 9.2, with $w = 1$ and first 180 lags. Training set: 01.03.10–31.05.10, testing set: 01.06.10–31.08.10.

9.4 Simulation

We recall that our two-dimensional model has the form

$$\begin{cases} X(t) = \int_{-\infty}^{t} g_X(t-s) dW^X(s), \\ Y(t) = \int_{-\infty}^{t} g_Y(t-s) dW^Y(s), \\ dW^X(s)\, dW^Y(s) = \rho dt, \end{cases}$$

where the kernels g_X and g_Y are given by Eq. (8.2). For many applications (beyond the scope of this thesis) the practitioners need to simulate from the model. Following [Bennedsen et al., 2017], we propose a *hybrid scheme*, which allows us to generate paths of the desired process. Let $n \in \mathbb{N}$. Then

$$\begin{aligned} X(t) &= \sum_{k=1}^{\infty} \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} g_X(t-s) dW^X(s), \\ Y(t) &= \sum_{k=1}^{\infty} \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} g_Y(t-s) dW^Y(s). \end{aligned}$$

We know that g_X and g_Y satisfy (C1), i.e.

$$\begin{aligned} g_X(t) &= t^{\alpha_X} L_g^X(t) \\ g_Y(t) &= t^{\alpha_Y} L_g^Y(t), \end{aligned}$$

for $\alpha_X, \alpha_Y \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$ and continuously differentiable functions $L_g^X(t)$ and $L_g^Y(t)$, slowly varying at zero and bounded away from zero. Therefore we can approximate

- $g_X(t-s) \approx (t-s)^{\alpha_X} L_g^X\left(\frac{k}{n}\right)$ for $t-s \in \left[\frac{k-1}{n}, \frac{k}{n}\right]$ and k "small";
- $g_X(t-s) \approx g_X\left(\frac{b_k}{n}\right)$ for $b_k \in [k-1, k]$ and k "large" (at least bigger than 2).

This means that for $\kappa = 1, 2, \ldots$

$$\begin{aligned} X(t) &= \sum_{k=1}^{\infty} \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} g_X(t-s) dW^X(s) \\ &\approx \sum_{k=1}^{\kappa} L_g^X\left(\frac{k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} (t-s)^{\alpha_X} dW^X(s) \\ &+ \sum_{k=\kappa+1}^{\infty} g_X\left(\frac{b_k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} dW^X(s). \end{aligned} \tag{9.12}$$

If $\kappa = 0$, we interpret the first sum on the right-hand side of Eq. (9.12) as zero and require $b_1 \in (0, 1]$. We define the cut-off point for the second sum as $N_n \geq \kappa + 1$. We approximate g_Y analogously. Now we can simulate from the model defined in Chapter 9 as follows.

1. We define constants $\kappa \in \mathbb{N}$ and $N_n \geq \kappa + 1$. We define also a grid $G_n = \{t, t - \frac{1}{n}, t - \frac{2}{n}, \ldots\}$ for some $n \in \mathbb{N}$.
2. We choose a sequence $\{b_k\}_{k \geq \kappa+1}$ such that $b_k \in [k-1, k] \setminus \{0\}$ for all $k \geq \kappa + 1$.
3. We simulate paths of correlated Brownian motions W^X and W^Y with the correlation coefficient ρ.
4. We simulate

$$\begin{aligned} &X^{\mathrm{sim}}(t) \\ &= \sum_{k=1}^{\kappa} L_g^X\left(\frac{k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} (t-s)^{\alpha_X} \left(W^X\left(t - \frac{k}{n} + \frac{1}{n}\right) - W^X\left(t - \frac{k}{n}\right)\right) \\ &+ \sum_{k=\kappa+1}^{\infty} g_X\left(\frac{b_k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} \left(W^X\left(t - \frac{k}{n} + \frac{1}{n}\right) - W^X\left(t - \frac{k}{n}\right)\right), \\ &Y^{\mathrm{sim}}(t) \\ &= \sum_{k=1}^{\kappa} L_g^Y\left(\frac{k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} (t-s)^{\alpha_Y} \left(W^Y\left(t - \frac{k}{n} + \frac{1}{n}\right) - W^Y\left(t - \frac{k}{n}\right)\right) \\ &+ \sum_{k=\kappa+1}^{\infty} g_Y\left(\frac{b_k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} \left(W^Y\left(t - \frac{k}{n} + \frac{1}{n}\right) - W^Y\left(t - \frac{k}{n}\right)\right). \end{aligned}$$

5. We compute

$$S^{\text{sim}}(t) = \sqrt{X^{\text{sim}}(t)^2 + Y^{\text{sim}}(t)^2}.$$

For more details and theoretical properties of the scheme we refer the reader to [Bennedsen et al., 2017].

9.5 Forecasting wind speed directly

One can wonder if the two-dimensional approach is worth pursuing. In other words, does it give better results than forecasting wind speed directly, using a one-dimensional model? We investigate this question by modelling and forecasting the wind speed with the following model.

$$S(t) = \int_{-\infty}^{t} g_S(t-s)dW(s),$$

where for $t > 0$

$$g_S(t) = \sigma_S t^{\alpha_S} \left(1 + \frac{t}{60}\right)^{\beta_S} \exp\left(-\frac{t}{1440}\right),$$

with $\sigma_S > 0$, $\alpha_S > -\frac{1}{2}$ and $\beta_S \in \mathbb{R}$. We calibrate this model using the regression-based approach described in Subsection 8.2.1.2. Figure 9.8 compares the forecasting performance of this direct wind speed model with the persistence forecast and forecasts obtained from two horizontal components as described in Section 9.3. We train the models on 01.12.09–28.02.10, with $w = 1$ and first 180 lags. We test them on 01.03.10–31.03.10.
We can see that while for very short horizons all methods' performances are comparable, for horizons longer than about 20 minutes both approximations of wind speed forecasts based on horizontal components outperform direct forecasts. We expect we would notice an even bigger difference in performances if we considered probabilistic rather than just point forecasts. Finally, in some locations, such as valleys, wind directions and speeds tend to be highly correlated, which suggests the component split as opposed to the direct approach ignoring any information about the directions.

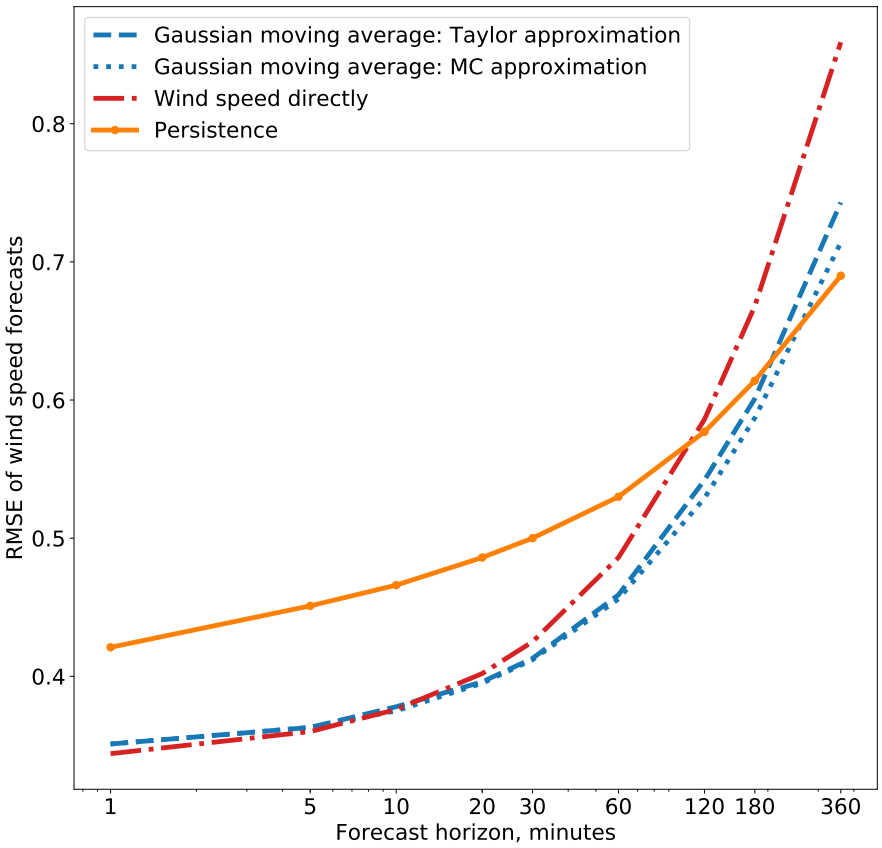


Figure 9.8: Root mean square errors of forecasted wind speed S, approximated with a Taylor series and Monte Carlo simulation, compared to the persistence forecast and direct wind speed forecast. Models calibrated with the method described in Section 9.2, with $w = 1$ and first 180 lags. Training set: 01.12.09–28.02.10, testing set: 01.03.10–31.03.10.

9.6 Summary and outlook

This chapter extended the results of Chapter 8, where we proposed a model for horizontal components of wind speed based on Gaussian moving average processes with a novel triple-scale kernel. We suggested a way to combine these one-dimensional models into a two-dimensional wind speed model. First, we defined bivariate Gaussian semistationary processes and studied their properties. To model the wind speed we started from the assumption of uncorrelated horizontal components and extended it to a model with a constant correlation coefficient between the components. We developed a model estimation method which allows to put more weight on either marginal distributions or the correlation structure, depending on the particular application.
After fitting the model to the data, we approximated wind speed forecasts in two ways: using the Taylor series approximation and Monte Carlo method. Both prediction methods significantly outperformed persistence forecasts for short time horizons, according to the root mean square error. We also checked how the forecasts resulting from the component split compare to predicting wind speed directly with a one-dimensional Gaussian moving average model with the triple-scale kernel. As expected, modelling horizontal components separately results in better forecasts for horizons exceeding about 20 minutes. To round the chapter off we recalled a method of simulating from our model.
Further studies should focus on identifying the best way of modelling the dependence structure between horizontal components. One could consider replacing the driving Brownian motion by a Lévy process, for example with a generalised hyperbolic marginal distribution. While distributions from this class provide a better fit to the component data than the Gaussian distribution, the Gaussian framework makes modelling and forecasting easier, so one would have to carefully weigh benefits and drawbacks. Another approach would be to replace the Brownian semistationary framework by multivariate trawl models, processes suitable for modelling a variety of time series, for instance in economics ([Barndorff-Nielsen et al., 2014]) or environmental sciences ([Noven et al., 2018]). This approach would involve establishing suitable forecasting methods tailored to trawl processes. Finally, autoregressive fractionally integrated moving average processes (ARFIMA) and their

continuous-time counterparts ([Marquardt, 2007]) are good candidates for modelling time series with long memory, for which the long-run mean decays slower than exponentially.

10

Conclusions and further studies

10.1 Conclusions

In this thesis we used stochastic modelling and statistical inference to answer two crucial questions posed by energy providers:

- How does the level of wind energy generation impact electricity prices?
- Can we produce accurate short-term wind speed forecasts without resorting to computationally expensive methods and the access to elaborate physical models?

In Part I we tackled the first problem. To this end, we extended the model of electricity spot prices proposed by [Benth et al., 2014] to include stochastic volatility present increasingly often in markets with high participation of renewable sources, as well as the information about wind energy generation. We also suggested a way to model wind-related variables. The latter extension not only provides a valuable economic interpretation but also, according to our empirical studies, improves the model fit. Finally, we proposed a joint model of electricity prices and wind energy generation particularly suitable

for studying the risks faced by energy distributors and traders.
The second question we answered positively in Part II. We found that while modelling wind speed directly did not give satisfactory results, splitting it into two horizontal components allowed us to produce accurate short-term forecasts. We successfully applied moving average Gaussian processes – univariate to model horizontal components and correlated bivariate for the wind speed – with a novel triple-scale kernel able to capture the complex behaviour of wind speed autocorrelation structure. After developing and testing a variety of estimation procedures, we managed to produce short-term forecasts which outperformed not only standard persistence forecasts but also forecasts without the component split.

10.2 Further studies

Our answers encourage even more research questions. We already concluded Chapter 4, Chapter 5, Chapter 8 and Chapter 9 with suggestions of studies extending presented results. In the final section of the thesis we would like to go one step further and see how our work fits into the broader topic of energy modelling.
Our research focused on temporal models, which could serve as a building block for spatio-temporal models. In particular, one could extend the wind speed modelling and forecasting framework proposed in Part II to multiple locations within a region or even a country. This would translate into the prediction of total wind energy production, thanks to its relationship with wind speed expressed as the power curve ([Hering and Genton, 2010]).
Particularly interesting objects useful in spatio-temporal modelling are ambit fields ([Barndorff-Nielsen et al., 2018]), which encompass also purely temporal processes described in this thesis, such as CARMA (Appendix C) or Gaussian moving average processes (Eq. (8.1)). Applications of such spatio-temporal processes range from modelling sea surface temperature anomalies ([Nguyen and Veraart, 2017]) to tumour growth ([Jónsdóttir et al., 2008]) to turbulence ([Barndorff-Nielsen and Schmiegel, 2004]). Given the usually strong spatial dependence of the wind ([Bardal and Sætran, 2016]), we expect that a similar

approach, carefully applied, would improve the models and forecasts proposed in Part II.

One can easily see how the forecasts of total wind energy generation in a country would fit directly into our model from Chapter 4. Ideally, the spatio-temporal model would have enough flexibility to incorporate whole energy networks encompassing multiple countries. This would fulfil the growing need of energy producers and traders, given the increasing volume of cross-border energy exchanges.

We firmly believe that such mathematical and statistical studies on wind energy and electricity prices will encourage new investments in renewable energy sources. Before our chance to stop climate change is gone with the wind.

A

Generalised hyperbolic distributions

In this thesis we often consider distributions from the generalised hyperbolic (GH) class. Their properties such as fat tails and skewness make them suitable for our applications. Most importantly, they are infinitely divisible, which in particular means that they can generate a Lévy process (see Appendix B) $L(t)$ such that $L(1)$ has a generalised hyperbolic density.
[Lüthi and Breymann, 2016] provided a concise introduction to these distributions and developed an `R` package `ghyp`, which came in useful many times during our research.

A.1 Definitions and properties

We begin by defining the building block of these Gaussian mixture distributions: the generalised inverse Gaussian distribution (GIG).

Definition A.1 (**Generalised inverse Gaussian distribution**)**.** The density of a *generalised inverse Gaussian* variable $W \sim GIG(\lambda, \chi, \psi)$ with parameters satisfying one of the following

- $\chi > 0, \psi \geq 0, \lambda < 0,$

- $\chi > 0, \psi > 0, \lambda = 0$,
- $\chi \geq 0, \psi > 0, \lambda > 0$,

is given by

$$f_{\text{GIG}}(x) = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{x^{\lambda-1}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} \exp\left(-\frac{1}{2}\left(\frac{\chi}{x} + \psi x\right)\right).$$

Now we are ready to define generalised hyperbolic distributions as Gaussian mixtures.

Definition A.2 (**Generalised hyperbolic distribution**)**.** A random vector X follows a *univariate generalised hyperbolic (GH) distribution* if

$$X \overset{law}{=} \mu + W\gamma + \sqrt{W}\sigma Z, \tag{A.1}$$

where $Z \sim N(0,1)$, $\sigma, \mu, \gamma \in \mathbb{R}$ and W is a scalar-valued random variable, independent of Z, whose distribution is generalised inverse Gaussian: $GIG(\lambda, \chi, \psi)$.

We interpret the parameters of generalised hyperbolic distribution as follows.

- μ corresponds to the location.
- $\Sigma = \sigma^2$ measures the dispersion.
- λ, χ and ψ determine the shape, i.e. the relationship between the tails and the center: as a rule of thumb, their large values indicate that the distribution resembles a Gaussian distribution.
- γ indicates the skewness: $\gamma = 0$ means that the distribution is symmetric.

We also observe that

$$X|W = w \sim N\left(\mu + w\gamma, w\Sigma\right).$$

We recall one more desirable property of this distribution class.

Proposition A.0.1 ([Lüthi and Breymann, 2016, Proposition 1])**.** *The generalised hyperbolic class is closed under linear transformations: if*

$$X \sim GH\left(\lambda, \chi, \psi, \mu, \Sigma, \gamma\right)$$

and

$$Y = aX + b$$

for $a, b \in \mathbb{R}$*, then*

$$Y \sim GH\left(\lambda, \chi, \psi, a\mu + b, a^2\Sigma, a\gamma\right).$$

A.2 Special cases of generalised hyperbolic distributions

In the class of generalised hyperbolic distributions we distinguish special subclasses.

- *Hyperbolic (hyp)* for $\lambda = 1$.
- *Normal inverse Gaussian (NIG)* for $\lambda = \frac{1}{2}$.
- *Variance gamma (VG)* for $\chi = 0$ and $\lambda > 0$ (a limiting case).
- *Generalised hyperbolic Student's-t* (often called *Student's-t*) for $\psi = 0$ and $\lambda < 0$.

In our research we use the `R` package `ghyp` ([Lüthi and Breymann, 2016]) to fit 11 different distributions to the increments of the Lévy process of interest and rank them according to the Akaike information criterion (AIC). Namely, we compare symmetric and asymmetric versions of hyp, NIG, VG, Student's-t and GH distributions, as well as a Gaussian distribution, to find the best fitting one.

A.3 Parametrisations

The $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ parametrisation has alternatives; the `R` package `ghyp` ([Lüthi and Breymann, 2016]) lets the user choose between three options. The other two are $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma,)$ and $(\lambda, \alpha, \mu, \Delta, \delta, \beta)$ parametrisations. [Lüthi and Breymann, 2016] provided ranges of parameters for special cases listed in Section A.2 as well as explicit functions switching between parametrisations. In this thesis we report the results using different parametrisations, depending on the application.

B

Lévy processes

For basic introductions to Lévy processes we suggest [Papapantoleon, 2008], [Barndorff-Nielsen and Shephard, 2012] or [Kyprianou, 2014]. More advanced readers could consult [Tankov and Cont, 2015], who not only provided clear explanations but also most proofs omitted in this chapter.
To study Lévy processes – stochastic processes able to accommodate jumps – we need to define the following special class of functions. We assume a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{R}}, \mathbb{P})$ satisfying the usual conditions: right-continuity of the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}}$ and completeness ($\mathcal{F}_0$ contains all $\mathbb{P}$-null sets).

Definition B.1 (**Cádlág**)**.** A function defined on (the subset of) the real line is *cádlág* (French: *continue á droite, limite á gauche*) or *RCLL* (*right-continuous with left limits*) if it is right-continuous everywhere and has left limits everywhere.

Now we are ready to define one of the most fundamental concepts of this thesis.

Definition B.2 (**Lévy process**)**.** We call a cádlág, adapted, real valued stochastic process $L = \{L(t)\}_{t\geq 0}$ a *Lévy process* if:

1. $L(0) = 0$ almost surely;
2. L has independent increments, i.e. $L(t) - L(s)$ is independent of $\{\mathcal{F}_s\}$ for all $0 \leq s < t < \infty$;
3. L has stationary increments, i.e. $L(t) - L(s)$ has the same distribution as $L(t-s)$ for all $0 \leq s < t < \infty$;
4. L is continuous in probability, i.e. $\lim_{s \to t} \mathbb{P}\left(L(t) - L(s) > \varepsilon\right) = 0$ for all $0 \leq s \leq t < \infty$ and $\varepsilon < 0$.

Some examples of Lévy processes include:

- Brownian motion: the only non-deterministic Lévy process with continuous paths.
- Poisson processes and compound Poisson processes.
- Stable processes.
- Lévy processes with generalised hyperbolic increments (see Appendix A).

If $L(t)$ is a Lévy processes, for $z \in \mathbb{R}$ and $t \geq 0$ its *characteristic function* has the form

$$\phi(t; z) = \mathbb{E}\left[izL(t)\right] = \exp\left(t\xi(z)\right),$$

where

$$\xi(z) = izb - \frac{z^2 c}{2} + \int_{-\infty}^{\infty} \left(e^{izx} - 1 - izx\mathbb{1}_{\{|x|<1\}}\right) \nu(dx),$$

$b \in \mathbb{R}$ is the drift, $c \geq 0$ the diffusion coefficient and ν a Lévy measure on the Borel subsets of $\mathbb{R}\backslash\{0\}$ such that

$$\int_0^{\infty} \min\left(x^2, 1\right) \nu(dx) < \infty.$$

We call $\xi(z)$ the *characteristic component* of $L(t)$.

C

Lévy-driven CARMA processes

We base this chapter on [García et al., 2011] and references therein, who provided an excellent introduction to the continuous version of widely used autoregressive moving average processes (ARMA).

C.1 ARMA models

[Brockwell and Davis, 2002] provided a detailed introduction to time series and forecasting, including autoregressive moving average (ARMA) processes.

Definition C.1 (ARMA(p,q) process). We call a time series $\{X_t\}_{t\in\mathbb{Z}}$ an autoregressive moving average process of order (p,q) or *ARMA(p,q)* if for all $t \in \mathbb{Z}$

$$X_t - \sum_{m=1}^{p} \phi_m X_{t-m} = \varepsilon_t + \sum_{n=1}^{q} \theta_n \varepsilon_{t-n},$$

where for all $t \in \mathbb{Z}$ and for a constant $\sigma > 0$ it holds that $\varepsilon_t \sim WN(0, \sigma^2)$ and polynomials $(1 - \sum_{m=1}^{p} \phi_m z^m)$, $(1 + \sum_{n=1}^{q} \theta_n z^n)$ have no common roots. We recall that $WN(0, \sigma^2)$ denotes white noise, i.e. a stationary sequence of uncorrelated random variables with mean zero and variance σ^2.

C.2 Definitions and properties

Definition C.2 (**CARMA process**). Let $L(t)$ be a Lévy process and $0 \leq q < p$. We define the *L-driven CARMA(p,q) process* with parameters $[a_1, \ldots, a_p; b_1, \cdots, b_q]$ as a stationary solution of the system of formal stochastic differential equations

$$a(D)Y(t) = b(D)DL(t), \tag{C.1}$$

where D denotes "differentiation" with respect to t,

$$a(z) := z^p + a_1 z^{p-1} + \cdots + a_p, \tag{C.2}$$

$$b(z) := b_0 + b_1 z + \cdots + b_{p-1}, \tag{C.3}$$

$a(z)$ and $b(z)$ have no common roots, $b_q = 1$ and $b_j = 0$ for $q < j < p$. Since we cannot define the derivative $DL(t)$ in the usual sense, we interpret Eq. (C.1) as the state-space representation, i.e.

$$Y(t) = \boldsymbol{b}^T \boldsymbol{X}(t), \tag{C.4}$$

$$d\boldsymbol{X}(t) = \boldsymbol{A}\boldsymbol{X}(t)dt + \boldsymbol{e}dL(t), \tag{C.5}$$

where

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{pmatrix}, \tag{C.6}$$

$$\boldsymbol{e} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-2} \\ b_{p-1} \end{pmatrix}. \tag{C.7}$$

Following [Brockwell et al., 2011], we assume that $\boldsymbol{X}(0)$ is independent of $\{L(t), t \geq 0\}$, all eigenvalues of $\boldsymbol{A}$ (easily calculated as roots of the autoregressive polynomial $a(z)$) have negative real parts and

$$\boldsymbol{X}(0) \stackrel{law}{=} \int_0^\infty e^{\boldsymbol{A}(t-u)}\boldsymbol{e}dL(u). \tag{C.8}$$

Then $\{\boldsymbol{X}(t), t \geq 0\}$ is a strictly stationary solution of Eq. (C.5).
In order to define CARMA(p,q) on the real line, we can define a Lévy process $\{M(t), t \geq 0\}$ such that $L(t)$ and $M(t)$ are iid. Then we extend $L(t)$ for the whole real line as

$$\tilde{L}(t) := L(t)\mathbb{1}_{[0,\infty)}(t) - M(-t-)\mathbb{1}_{(-\infty,0]}(t), \quad t \in \mathbb{R}. \tag{C.9}$$

In the remaining part of this chapter we will denote $\tilde{L}$ by L and call it the *background driving Lévy process* of $Y(t)$. Now we are ready to provide the necessary and sufficient conditions for the existence of a covariance stationary solution of Eq. (C.5).

Proposition C.0.1 ([Brockwell et al., 2011, Proposition 1])**.** *Eq. (C.5) has a covariance stationary solution $\boldsymbol{X}$ such that $\boldsymbol{X}(t)$ is independent of $\{L(s) - L(t)\}_{s>t}$ for all $t \in \mathbb{R}$ if and only if the real parts of the roots of $a(z) = 0$ are negative. This solution has the form*

$$\boldsymbol{X}(t) = \int_{-\infty}^t e^{\boldsymbol{A}(t-u)}\boldsymbol{e}dL(u) \stackrel{law}{=} \int_0^\infty e^{\boldsymbol{A}u}\boldsymbol{e}dL(u).$$

This leads us to an alternative definition of the object crucial for this thesis: the continuous autoregressive moving average process (CARMA).

Definition C.3 (**CARMA(p,q) process**)**.** Let us assume that L is a Lévy process, the roots of $a(z) = 0$ are distinct and their real parts are negative. Then we define the *L-driven CARMA(p,q) process* with parameters $a_1, \ldots, a_p; b_1, \ldots, b_q$ as the process satisfying

$$Y(t) = \boldsymbol{b}^T\boldsymbol{X}(t) = \int_{-\infty}^\infty g(t-u)dL(u)$$

with

$$g(t) = \boldsymbol{b}^T e^{\boldsymbol{A}t} \boldsymbol{e} \mathbb{1}_{[0,\infty)}(t)$$

called the *kernel of the CARMA process* $\{Y(t)\}$. Y is a *causal* function of L because $Y(t)$ does not depend on $\{L(s) - L(t), s \geq t\}$ for all t.

To assure the existence and uniqueness of our CARMA processes, we need to state the following additional assumptions.

Assumption C.1. Unless stated otherwise, throughout the whole thesis we assume that roots of $a(z) = 0$ are distinct and their real parts are negative. Furthermore, $a(z)$ and $b(z)$ have no common roots.

CARMA processes have an intuitive and very useful representation as a sum of Lévy-driven Ornstein-Uhlenbeck processes, described in more detail in Appendix D. This observation helps us estimate the parameters of these objects as well as simulate them.

Proposition C.0.2 ([Brockwell et al., 2011, Proposition 2])**.** *Under Assumption C.1, any CARMA(p,q) process $Y(t)$ can be expressed as a sum of dependent and possibly complex-valued CAR(1) processes, i.e.*

$$Y(t) = \sum_{r=1}^{p} Y^{(r)}(t), \tag{C.10}$$

where for $r = 1, \ldots, p$

$$Y^{(r)}(t) = \int_{-\infty}^{t} \alpha_r e^{\lambda_r (t-u)} dL(u), \tag{C.11}$$

$$\alpha_r = \frac{b(\lambda_r)}{a'(\lambda_r)}, \tag{C.12}$$

with $a'(\cdot)$ denoting the derivative of $a(\cdot)$.

In this thesis we focus on one particular example: the CARMA(2,1) process. Therefore we state some basic facts about this object.

Example C.1 (CARMA(2,1))**.** The Lévy-driven CARMA(2,1) process is defined as the strictly stationary solution to the system of equations

$$(D^2 + a_1 D + a_2)Y(t) = (b_0 + D)DL(t), \tag{C.13}$$

where $t \in \mathbb{R}$, $b(z) = b_0 + z$ and $a(z) = z^2 + a_1 z + a_2 = (z - \lambda_1)(z - \lambda_2)$, $\lambda_1 \neq \lambda_2$, $Re(\lambda_1) < 0$ and $Re(\lambda_2) < 0$. Furthermore, for $u \geq 0$ the kernel of $Y(t)$ equals

$$g(u) = \alpha_1 e^{\lambda_1 u} + \alpha_2 e^{\lambda_2 u}, \tag{C.14}$$

where $\alpha_1 = \frac{b_0+\lambda_1}{\lambda_1-\lambda_2}$ and $\alpha_2 = \frac{b_0+\lambda_2}{\lambda_2-\lambda_1}$. Therefore we can represent this process as

$$Y(t) = \alpha_1 \int_{-\infty}^{t} e^{\lambda_1(t-u)} dL(u) + \alpha_2 \int_{-\infty}^{t} e^{\lambda_2(t-u)} dL(u). \tag{C.15}$$

C.3 The sampled CARMA process

In practice we observe any continuous process $Y(t)$ only in discrete time. Precisely, we denote by $\{Y_n := Y(nh), n = 0, 1, \ldots, N\}$ the sampled process, where N is the number of available observations and $h > 0$ is a small, fixed interval between the consecutive observations. The following proposition explores the relationship between the continuous and discrete processes, which prepares us for the estimation of CARMA(p,q) parameters.

Proposition C.0.3 ([Brockwell et al., 2011, Proposition 3])**.** *Under Assumption C.1 the following hold.*

1. *The sampled process $\{Y_n := Y(nh), n = 0, 1, \ldots, N\}$, with a fixed $h > 0$, can be represented as $Y_n = \sum_{r=1}^{p} Y_n^{(r)}$, where $n \in \mathbb{Z}$. For each $r = 1, \ldots, p$, the discrete-time process $\{Y_n^{(r)}\}$ is obtained by sampling the component CAR(1) process $\{Y^{(r)}(t)\}$ (cf. Appendix D) at spacing $h > 0$. As Y is strictly stationary,*

$$\forall n \in \mathbb{Z} \quad Y_n^{(r)} = e^{\lambda_r} Y_{n-1}^{(r)} + Z_n^{(r)}, \tag{C.16}$$

with the iid noise

$$\forall n \in \mathbb{Z} \quad Z_n^{(r)} = \alpha_r \int_{(n-1)h}^{nh} e^{\lambda_r (nh-u)} dL(u). \tag{C.17}$$

2. *The sampled process* $\{Y_n := Y(nh), n = 0, 1, \ldots, N\}$ *satisfies*

$$\phi(B) Y_n = \sum_{r=1}^{p} V_{n-r+1}^r =: U_n, \tag{C.18}$$

where

$$\phi(z) := \prod_{r=1}^{p} \left(1 - e^{\lambda_r h} z\right) = 1 - \sum_{r=1}^{p} \phi_r z^r \tag{C.19}$$

and B *denotes the backshift operator, i.e.* $B^j Y_n := Y_{n-j}$. *For each* $r = 1, \ldots, p$, *we define the iid sequence* $\{V_n^{(r)}\}$ *by*

$$V_n^{(r)} := \int_{(n-1)h}^{nh} \sum_{k=1}^{p} \alpha_k \left(e^{(r-1)h\lambda_k} - \sum_{j=1}^{r-1} \phi_j e^{(r-1-j)h\lambda_k} \right) \times e^{(nh-u)\lambda_k} dL(u). \tag{C.20}$$

3. *We can represent the right-hand side of Eq. (C.18) as an invertible moving average*

$$\theta(B) W_n := W_n + \theta_1 W_{n-1} + \cdots + \theta_{p-1} W_{n-p+1}, \tag{C.21}$$

where $\{W_n\}$ *is a sequence of white noise (possibly not iid) and* $\theta_1, \ldots, \theta_q$ *are moving average constants depending on the CARMA process. Therefore* $\{Y_n\}$ *can be represented as a weak ARMA(p,p-1) process (an ARMA(p,p-1) process allowing for not iid driving white noise) such that*

$$\phi(B) Y_n = \theta(B) W_n \tag{C.22}$$

and

$$W_n = \theta(B)^{-1} \sum_{r=1}^{p} V_{n-r+1}^p. \tag{C.23}$$

C.4 Estimation

Now we are ready to estimate the CARMA parameters, following the algorithm described by [García et al., 2011]. [Brockwell et al., 2011] showed that for a fixed sampling interval $h > 0$ the mean corrected sampled CARMA(p,q) process is a weak ARMA(p,p-1) process, so we can estimate ARMA(p,p-1) parameters and map them to the continuous setting as outlined below. In the thesis we are particularly interested in CARMA(2,1), so for brevity we specify the estimation procedure for this particular process. We slightly abuse the notation by denoting variables and their estimates with the same symbols; the meaning should be clear from the context.

1. We estimate ARMA(2,1) parameters $\boldsymbol{\beta} = (\phi_1, \phi_2, \theta)^T$, using the quasi-maximum likelihood approach (alternatively least squares).

2. Eq. (C.18) has the form

$$Y_n - \phi_1 Y_{n-1} - \phi_2 Y_{n-2} = \left(1 - e^{\lambda_1 h} B\right)\left(1 - e^{\lambda_2 h} B\right) Y_n. \tag{C.24}$$

   By multiplying through and matching coefficients, we obtain

$$\phi_1 = e^{\lambda_1 h} + e^{\lambda_2 h}, \quad \phi_2 = -e^{(\lambda_1 + \lambda_2) h}. \tag{C.25}$$

   This gives us a nonlinear system of two equations for the estimators of λ_1 and λ_2, whose solutions are

$$\lambda_1 = \log\left(\frac{\phi_1}{2} + \sqrt{\left(\frac{\phi_1}{2}\right)^2 + \phi_2^2}\right), \tag{C.26}$$

$$\lambda_2 = \log\left(\frac{\phi_1}{2} - \sqrt{\left(\frac{\phi_1}{2}\right)^2 + \phi_2^2}\right). \tag{C.27}$$

   From there we immediately calculate $a_1 = -(\lambda_1 + \lambda_2)$ and $a_2 = \lambda_1 \lambda_2$.

   Remark C.1. We implicitly assume that $\phi_1 > 0$ and $\phi_2 < 0$, so this method is not suitable for some CARMA processes. Alternative estima-

tion methods include quasi-maximum likelihood implemented in the `R` package `yuima` ([Iacus and Mercuri, 2015]).

3. The right-hand side of Eq. (C.18) implies the form of the autocovariance of the process $\phi(B)Y_n$:

$$\forall s \in \mathbb{R} \quad \gamma_U(s) = \operatorname{cov}\left(\phi(B)Y_n, \phi(B)Y_{n-s}\right). \tag{C.28}$$

Furthermore, using Corollary 3 by [Barndorff-Nielsen et al., 2013], for all $s \in \mathbb{R}$ we calculate the autocovariance of $Y(t)$

$$\gamma_Y(s) = \operatorname{Cov}\left(Y(s), Y(0)\right) = \int_0^\infty g(x)g(x+s)dx = w_1 e^{\lambda_1 s} + w_2 e^{\lambda_2 s}, \tag{C.29}$$

where

$$w_1 = \frac{\alpha_1^2\lambda_1\lambda_2 + {\alpha_1}^2\lambda_2^2 + 2\lambda_1\lambda_2\alpha_1\alpha_2}{2\lambda_1\lambda_2\left(\lambda_1+\lambda_2\right)}, \tag{C.30}$$

$$w_2 = \frac{\alpha_2^2\lambda_1\lambda_2 + {\alpha_2}^2\lambda_1^2 + 2\lambda_1\lambda_2\alpha_1\alpha_2}{2\lambda_1\lambda_2\left(\lambda_1+\lambda_2\right)}. \tag{C.31}$$

Thus for all $s \in \mathbb{R}$ the autocorrelation of $Y(t)$ equals

$$\delta_Y(s) = \frac{\gamma_Y(s)}{\gamma_Y(0)} = \frac{w_1 e^{\lambda_1 s} + w_2 e^{\lambda_2 s}}{w_1 + w_2}. \tag{C.32}$$

For CARMA(2,1) we can rewrite Eq. (C.28) as

$$\begin{aligned}
\gamma_U(0) &= (1+\phi_1^2+\phi_2^2)\gamma_Y(0) + (2\phi_2\phi_2 - 2\phi_1)\gamma_Y(1) - 2\phi_2\gamma_Y(2),\\
\gamma_U(1) &= -\phi_2\gamma_Y(3) + \phi_1(\phi_2-1)\gamma_Y(2)\\
&\quad + (1+\phi_1^2+\phi_2^2-\phi_2)\gamma_Y(1) + \phi_1(\phi_2-1)\gamma_Y(0),
\end{aligned} \tag{C.33}$$

where we use explicit formulae for $\gamma_Y(\cdot)$ given by Eq. (C.29). Since they depend on a_0, a_1 and b_0, we plug in the estimates of the first two parameters.
On the other hand, the autocorrelation function at the first lag of

a moving average process with coefficient θ can be expressed as

$$\delta_U(1) = \frac{\gamma_U(1)}{\gamma_U(0)} = \frac{\theta}{1+\theta^2}. \tag{C.34}$$

Now we can replace the left-hand side of Eq. (C.34) by expressions from Eq. (C.33) to get a non-linear equation for b_0, which we solve numerically.

4. Having estimated the parameters of CARMA(2,1), we need to recover the background driving Lévy process L(t), using results by [Brockwell et al., 2011, Section 5]:

$$X^{(0)}(t) = X^{(0)}(0)e^{-b_0 t} + \int_0^t e^{-b_0(t-s)}Y(s)ds, \tag{C.35}$$

$$X^{(1)}(t) = DX^{(0)}(t) = -b_0 X^{(0)}(t) + Y(t). \tag{C.36}$$

The canonical state vector $\mathbf{Y}(t)$ is given by

$$\begin{pmatrix} Y^{(1)}(t) \\ Y^{(2)}(t) \end{pmatrix} = \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} -\lambda_2(b_0+\lambda_1) & (b_0+\lambda_1) \\ \lambda_1(b_0+\lambda_2) & -b_0+\lambda_2 \end{pmatrix} \begin{pmatrix} X^{(0)}(t) \\ X^{(1)}(t) \end{pmatrix}. \tag{C.37}$$

To recover the background driving Lévy process L(t) we can choose one of two equations, either with $r = 1$ or $r = 2$:

$$L(t) = \frac{1}{\alpha_r}\left[Y^{(r)}(t) - Y^{(r)}(0) - \lambda_r \int_0^t Y^{(r)}(s)ds\right]. \tag{C.38}$$

We approximate the integral using the trapezoidal rule. [Brockwell et al., 2011] recommend choosing r such that $|\lambda_r|$ is minimal, which minimizes the contribution of $\lambda_r \int_0^t Y^{(r)}(s)ds$ compared to $Y^{(r)}(t) - Y^{(r)}(0)$.

C.5 Simulation

To simulate any CARMA(p,q) process, we use its representation as a sum of CAR(1) processes stated in Proposition C.0.2. For brevity we restrict our attention to simulating CARMA(2,1) process $Y(t)$ with parameters a_1, a_2

and b_0 and eigenvalues of the matrix $\boldsymbol{A}$: λ_1 and λ_2. We start from simulating two Ornstein-Uhlenbeck processes $Y^{(1)}(t)$ and $Y^{(2)}(t)$ with λ_1 and λ_2, driven by the same Lévy process (cf. Section D.4). Then we compute

$$Y(t) = \alpha_1 Y^{(1)}(t) + \alpha_2 Y^{(2)}(t),$$

where

$$\begin{aligned} \alpha_1 &= \frac{b_0 + \lambda_1}{\lambda_1 - \lambda_2} \\ \alpha_2 &= \frac{b_0 + \lambda_2}{\lambda_1 - \lambda_2}. \end{aligned}$$

D

Lévy-driven CAR(1) processes

One of the most important processes belonging to the CARMA class (Appendix C) is the Lévy-driven CAR(1) process, called also Lévy-driven Ornstein-Uhlenbeck process. For more details we refer the reader to [Brockwell et al., 2007].

D.1 Definitions and properties

Definition D.1 (**Lévy-driven CAR(1) process**)**.** Let $L(t)$ be a Lévy process. The *L-driven Ornstein-Uhlenbeck process* with the parameter $\lambda > 0$ is a strictly stationary solution of the stochastic differential equation

$$dY(t) = -\lambda Y(t)dt + dL(t), \quad t > 0.$$

As in Eq. (C.9) we extend $L(t)$ to the whole real line. If $\lambda > 0$, for all $t > s \in \mathbb{R}$ the process $Y(t)$ defined by

$$Y(t) = \int_{-\infty}^{t} e^{-\lambda(t-u)} d\tilde{L}(u)$$

is a strictly stationary solution to

$$dY(t) = -\lambda Y(t)dt + d\tilde{L}(t).$$

Again, we will denote $\tilde{L}$ by L and call it the *background driving Lévy process* of $Y(t)$.

D.2 The sampled CAR(1) process

In this section we point out the relationship between the Ornstein-Uhlenbeck process and the discrete autoregressive process of order one. We denote by $\{Y_n := Y(nh), n = 0, 1, \ldots, N\}$ the sampled CAR(1) process, where N is the number of available observations and $h > 0$ is a small, fixed interval between the consecutive observations. The sampled CAR(1) process is the discrete time autoregressive AR(1) process (cf. Section C.1)

$$\forall n \in \mathbb{Z} \quad Y_n = \phi Y_{n-1} + Z_n,$$

where

$$\phi = e^{-\lambda h}$$

and

$$\forall n \in \mathbb{Z} \quad Z_n = \int_{(n-1)h}^{nh} e^{\lambda(nh-u)} dL(u)$$

is the iid and positive noise.

D.3 Estimation

To estimate the parameters of a CAR(1) process $Y(t)$ we follow the procedure suggested by [Brockwell et al., 2007] and references therein. Again, we slightly abuse the notation by denoting variables and their estimates with the same symbols, the meaning of which is clear from the context. We proceed as follows.

1. We estimate the parameter ϕ of the sampled process Y_n, for example by maximising the likelihood function.

2. We compute
$$\lambda = -\log(\phi).$$

3. For all $n = 1, \dots N$ we define the increments of the driving process $L(t)$ on intervals $((n-1)h), nh]$ by
$$\Delta_n^h L := L_{nh} - L_{(n-1)h} = Y_{nh} - Y_{(n-1)h} + \lambda \int_{(n-1)h}^{nh} Y_u du.$$
We recover these increments using the trapezoidal rule
$$\Delta_n^h L \approx Y_n - Y_{n-1} + \frac{\lambda h}{2} \left(Y_n + Y_{n-1}\right).$$

D.4 Simulation

We simulate a CAR(1) process $Y(t)$ as follows.

1. We partition the interval $[0, T]$ into N equal intervals of length $h = \frac{T}{N}$.

2. For $i = 1, \dots, N$ we simulate the increments of the driving process $\Delta_n^h L$ following the appropriate centred distribution, for example from the generalised hyperbolic class (see Appendix A). If the mean μ of Lévy increments differs from zero, after simulating the increments we subtract the mean from the resulting time series.

3. We set the initial value of Y to 0.

4. For $n = 1, \dots, (N-1)$ we use the Euler-Maruyama method to compute
$$Y_{n+h} = Y_n + \lambda h Y_n + (L_{n+h} - L_n) + h\mu.$$

In practice we first simulate a CAR(1) process of length $2N$. We further discard the first half of the time series as a burn-in period. We finally resample

the resulting time series every $\frac{1}{h}$ points to obtain a process corresponding to the original data, if needed.

Bibliography

[Aïd, 2015] Aïd, R. (2015). *Electricity derivatives*. SpringerBriefs in Quantitative Finance. Springer International Publishing.

[Ak et al., 2016] Ak, R., Fink, O., and Zio, E. (2016). Two machine learning approaches for short-term wind speed time-series prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 27(8):17341747.

[Austrian Power Grid, 2019] Austrian Power Grid (2019). Generation forecast. https://www.apg.at/en/markt/Markttransparenz/erzeugung/erzeugungsprognose. Accessed: 20.02.2019.

[Bank for International Settlements, 1996] Bank for International Settlements (1996). Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements. https://www.bis.org/publ/bcbs22.htm.

[Bardal and Sætran, 2016] Bardal, L. M. and Sætran, L. R. (2016). Spatial correlation of atmospheric wind at scales relevant for large scale wind turbines. *Journal of Physics: Conference Series*, 753(3):0–7.

[Barndorff-Nielsen et al., 2018] Barndorff-Nielsen, O., Benth, F., and Veraart, A. (2018). *Ambit Stochastics*. Probability Theory and Stochastic Modelling. Springer International Publishing.

[Barndorff-Nielsen et al., 2013] Barndorff-Nielsen, O. E., Benth, F. E., and Veraart, A. E. D. (2013). Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. *Bernoulli*, 19(3):803–845.

[Barndorff-Nielsen et al., 2014] Barndorff-Nielsen, O. E., Lunde, A., Shephard, N., and Veraart, A. E. D. (2014). Integer-valued trawl processes: A

class of stationary infinitely divisible processes. *Scandinavian Journal of Statistics*, 41(3):693–724.

[Barndorff-Nielsen and Schmiegel, 2004] Barndorff-Nielsen, O. E. and Schmiegel, J. (2004). Lévy-based tempo-spatial modelling; with applications to turbulence. *Uspekhi Mat. Nauk*, 159.

[Barndorff-Nielsen and Shephard, 2012] Barndorff-Nielsen, O. E. and Shephard, N. (2012). Basics of Lévy processes. http://www.nuff.ox.ac.uk/economics/papers/2012/introlevy120608.pdf.

[Bennedsen, 2015] Bennedsen, M. (2015). Rough electricity: a new fractal multi-factor model of electricity spot prices. CREATES Research Papers 2015-42, Department of Economics and Business Economics, Aarhus University.

[Bennedsen et al., 2017] Bennedsen, M., Lunde, A., and Pakkanen, M. S. (2017). Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965.

[Benth et al., 2014] Benth, F. E., Klüppelberg, C., Müller, G., and Vos, L. (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energy Economics*, 44:392–406.

[Benth and Paraschiv, 2016] Benth, F. E. and Paraschiv, F. (2016). A Structural Model for Electricity Forward Prices. Working Papers on Finance 1611, University of St. Gallen, School of Finance.

[Benth and Benth, 2010] Benth, J. Š. and Benth, F. E. (2010). Analysis and modelling of wind speed in New York. *Journal of Applied Statistics*, 37:893–909.

[Borovkova and Geman, 2006] Borovkova, S. and Geman, H. (2006). Seasonal and stochastic effects in commodity forward curves. *Review of Derivatives Research*, 9(2):167–186.

[Brockwell and Davis, 2002] Brockwell, P. and Davis, R. (2002). *An Introduction to Time Series and Forecasting*. Springer-Verlag New York.

[Brockwell et al., 2007] Brockwell, P. J., Davis, R. A., and Yang, Y. (2007). Estimation for non-negative Lévy-driven Ornstein-Uhlenbeck processes. *Journal of Applied Probability*, 44(4):977989.

[Brockwell et al., 2011] Brockwell, P. J., Davis, R. A., and Yang, Y. (2011). Estimation for non-negative Lévy-driven CARMA processes. *Journal of Business & Economic Statistics*, 29(2):250–259.

[Carmona et al., 2013] Carmona, R., Coulon, M., and Schwarz, D. (2013). Electricity price modeling and asset valuation: A multi-fuel structural approach. *Mathematics and Financial Economics*, 7(2):167–202.

[Cartea et al., 2009] Cartea, Á., Figueroa, M. G., and Geman, H. (2009). Modelling electricity prices with forward looking capacity constraints. *Applied Mathematical Finance*, 16(2):103–122.

[Chaudhuri, 1996] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872.

[Czado, 2019] Czado, C. (2019). *Analyzing Dependent Data with Vine Copulas: A Practical Guide With R.* Springer International Publishing.

[Davis, 2006] Davis, R. A. (2006). *Encyclopedia of Environmetrics*, chapter Gaussian Process. John Wiley & Sons.

[Deschatre and Veraart, 2017] Deschatre, T. and Veraart, A. E. D. (2017). A joint model for electricity spot prices and wind penetration with dependence in the extremes. *SSRN: https://ssrn.com/abstract=3029318.*

[Drost, 2018] Drost, H. (2018). Philentropy: Information theory and distance quantification with R. *Journal of Open Source Software*, 3(26):765.

[Elberg and Hagspiel, 2015] Elberg, C. and Hagspiel, S. (2015). Spatial dependencies of wind power and interrelations with spot price dynamics. *European Journal of Operational Research*, 241(1):260–272.

[EPEX SPOT, 2019] EPEX SPOT (2019). Negative prices Q&A. https://www.epexspot.com/en/company-info/basics_of_the_power_market/negative_prices. Accessed: 24.10.2019.

[European Commission, 2007] European Commission (2007). Blackout of November 2006: important lessons to be drawn. Press Release. https://ec.europa.eu/commission/presscorner/detail/en/IP_07_110.

[European Commission, 2009] European Commission (2009). Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC. *Official Journal of the European Union*, L 140:39–85.

[European Commission, 2018a] European Commission (2018a). Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources. *Official Journal of the European Union*, L 328:92–209.

[European Commission, 2018b] European Commission (2018b). Study on the quality of electricity market data of transmission system operators, electricity supply disruptions, and their impact on the European electricity markets. Technical report, European Commission. https://ec.europa.eu/energy/sites/ener/files/documents/dg_ener_electricity_market_data_-_final_report_-_22032018.pdf.

[European Energy Exchange, 2017] European Energy Exchange (2017). Market data. https://www.eex.com/en/market-data/power/futures. Accessed: 10.11.2017.

[European Energy Exchange, 2018] European Energy Exchange (2018). Preparation for the split of the power market bidding zone between Germany and Austria. https://www.eex.com/blob/85382/d9fc314b4ced4b8c6b1102b3a6e8a6cf/20180904-customer-information---bidding-zone-split-data.pdf. Accessed: 16.09.2019.

[European Wind Energy Association, 2010] European Wind Energy Association (2010). Wind energy and electricity prices. Exploring the "merit order effect". http://www.ewea.org/fileadmin/files/library/publications/reports/MeritOrder.pdf.

[García et al., 2011] García, I., Klüppelberg, C., and Müller, G. (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Statistical Modelling*, 11(5):447–470.

[Giebel, 2003] Giebel, G. (2003). The state-of-the-art in short-term prediction of wind power: A literature overview. Technical report, Project ANEMOS (EU Framework 5).

[Gneiting et al., 2006] Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2006). Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time method. *Journal of the American Statistical Association*, 101(475):968–979.

[Haeffelin et al., 2005] Haeffelin, M., Barthès, L., Bock, O., Boitel, C., Bony, S., Bouniol, D., Chepfer, H., Chiriaco, M., Cuesta, J., Delanoë, J., Drobinski, P., Dufresne, J. L., Flamant, C., Grall, M., Hodzic, A., Hourdin, F., Lapouge, F., Lemaître, Y., Mathieu, A., Morille, Y., Naud, C., Noël, V., O'Hirok, W., Pelon, J., Pietras, C., Protat, A., Romand, B., Scialom, G., and Vautard, R. (2005). SIRTA, a ground-based atmospheric observatory for cloud and aerosol research. *Annales Geophysicae*, 23(2):253–275.

[Hall, 1996] Hall, A. R. (1996). *Generalized Method of Moments: Advanced Text in Econometrics*, volume 100-101. Oxford University Press.

[Hering and Genton, 2010] Hering, A. S. and Genton, M. G. (2010). Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, 105(489):92–104.

[Hida and Hitsuda, 1993] Hida, T. and Hitsuda, M. (1993). *Gaussian processes*, volume 120. American Mathematical Society.

[Hong, 2014] Hong, T. (2014). Load, demand, energy and power. http://blog.drhongtao.com/2014/09/load-demand-energy-power.html. Accessed: 01.12.2019.

[Hyndman and Khandakar, 2008] Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22.

[Iacus and Mercuri, 2015] Iacus, S. M. and Mercuri, L. (2015). Implementation of Lévy CARMA model in Yuima package. *Computational Statistics*, 30(4):1111–1141.

[International Renewable Energy Agency, 2018] International Renewable Energy Agency (2018). Renewable energy statistics 2018. https://irena.org/publications/2018/Jul/Renewable-Energy-Statistics-2018.

[Jónsdóttir et al., 2008] Jónsdóttir, K. Ý., Schmiegel, J., and Jensen, E. B. V. (2008). Lévy-based growth models. *Bernoulli*, 14(1):62–90.

[Jónsson et al., 2010] Jónsson, T., Pinson, P., and Madsen, H. (2010). On the market impact of wind energy forecasts. *Energy Economics*, 32(2):313–320.

[Ketterer, 2014] Ketterer, J. C. (2014). The impact of wind power generation on the electricity price in Germany. *Energy Economics*, 44:270–280.

[Kyprianou, 2014] Kyprianou, A. (2014). *Fluctuations of Lévy Processes with Applications: Introductory Lectures*. Universitext. Springer Berlin Heidelberg.

[Lang et al., 2019] Lang, M. N., Mayr, G. J., Stauffer, R., and Zeileis, A. (2019). Bivariate Gaussian models for wind vectors in a distributional regression framework. *Advances in Statistical Climatology, Meteorology and Oceanography*, 5(2):115–132.

[Lingohr and Müller, 2019] Lingohr, D. and Müller, G. (2019). Stochastic modeling of intraday photovoltaic power generation. *Energy Economics*, 81:175–186.

[Lüthi and Breymann, 2016] Lüthi, D. and Breymann, W. (2016). *ghyp: A Package on Generalized Hyperbolic Distribution and Its Special Cases*. R package version 1.5.7.

[Mandelbrot and Van Ness, 1968] Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437.

[Marquardt, 2007] Marquardt, T. (2007). Multivariate fractionally integrated CARMA processes. *Journal of Multivariate Analysis*, 98(9):1705–1725.

[Matyas et al., 1999] Matyas, L., MATYAS, L., Gourieroux, C., and Phillips, P. (1999). *Generalized Method of Moments Estimation*. Generalized Method of Moments Estimation. Cambridge University Press.

[McLloyd, 2013] McLloyd, F. (2013). Regelzonen mit Übertragungsnetzbetreiber in Deutschland. https://upload.wikimedia.org/wikipedia/commons/8/82/Regelzonen_mit_%C3%9Cbertragungsnetzbetreiber_in_Deutschland.png. Accessed: 04.09.2019. CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0).

[Met Office, 2019] Met Office (2019). Long range forecast. https://www.metoffice.gov.uk/weather/long-range-forecast. Accessed: 18.09.2019.

[Monahan, 2013] Monahan, A. H. (2013). The Gaussian statistical predictability of wind speeds. *Journal of Climate*, 26(15):55635577.

[Müller and Seibert, 2019] Müller, G. and Seibert, A. (2019). Bayesian estimation of stable CARMA spot models for electricity prices. *Energy Economics*, 78:267–277.

[Nguyen and Veraart, 2017] Nguyen, M. and Veraart, A. E. (2017). Modelling spatial heteroskedasticity by volatility modulated moving averages. *Spatial Statistics*, 20:148–190.

[Nicolosi and Fürsch, 2009] Nicolosi, M. and Fürsch, M. (2009). The impact of an increasing share of RES-E on the conventional power market – the example of Germany. *Zeitschrift für Energiewirtschaft*, 33(3):246–254.

[Noven et al., 2018] Noven, R. C., Veraart, A. E. D., and Gandy, A. (2018). A latent trawl process model for extreme values. *Journal of Energy Markets*, 11(3):1–24.

[Open Power System Data, 2019] Open Power System Data (2019). Data package time series. Version 2019-06-05. https://doi.org/10.25832/time_series/2019-06-05. (Primary data from various sources, for a complete list see URL).

[Papapantoleon, 2008] Papapantoleon, A. (2008). An introduction to Lévy processes with applications in finance. *arXiv: https://arxiv.org/pdf/0804.0482.pdf*.

[Pfaff, 2008] Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software*, 27(4).

[Pinson, 2012] Pinson, P. (2012). Adaptive calibration of (u,v)-wind ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138(666):12731284.

[Press et al., 1992] Press, W. H., Teukolsky, S. A., and Vetterling, W. T. (1992). *Numerical Recipes in C The Art of Scientific Computing. Second Edition*. Cambridge University Press.

[R Core Team, 2018] R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

[Rowińska et al., 2018] Rowińska, P. A., Veraart, A. E., and Gruet, P. (2018). A multifactor approach to modelling the impact of wind energy on electricity spot prices. *SSRN: https://ssrn.com/abstract=3110554*.

[Sato et al., 1999] Sato, K., Ken-Iti, S., Bollobas, B., Press, C. U., Katok, A., Fulton, W., Kirwan, F., and Sarnak, P. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

[Savitzky and Golay, 1964] Savitzky, A. and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):16271639.

[Schuhen et al., 2012] Schuhen, N., Thorarinsdottir, T. L., and Gneiting, T. (2012). Ensemble model output statistics for wind vectors. *Monthly Weather Review*, 140(10):3204–3219.

[Sergio and Ludermir, 2015] Sergio, A. T. and Ludermir, T. B. (2015). Deep learning for wind speed forecasting in northeastern region of Brazil. In *2015 Brazilian Conference on Intelligent Systems (BRACIS)*, pages 322–327.

[Sloughter et al., 2013] Sloughter, M. M., Gneiting, T., and Raftery, A. E. (2013). Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Monthly Weather Review*, 141(6):2107–2119.

[Tankov and Cont, 2015] Tankov, P. and Cont, R. (2015). *Financial Modelling with Jump Processes, Second Edition*. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis.

[Tao, 2011] Tao, T. (2011). *An Introduction to Measure Theory Vol. 126*. American Mathematical Society.

[ter Haar, 2010] ter Haar, R. (2010). On modelling the electricity futures curve.

[Trapletti and Hornik, 2018] Trapletti, A. and Hornik, K. (2018). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-46.

[United Nations, 2015] United Nations (2015). Multilateral Treaties Deposited with the Secretary-General. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement. Accessed: 30.11.2019.

[Venables and Ripley, 2002] Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition.

[Veraart, 2016] Veraart, A. E. D. (2016). Modelling the impact of wind power production on electricity prices by regime-switching Lévy semistationary processes. In Benth, F. E. and Di Nunno, G., editors, *Stochastics of Environmental and Financial Economics*, pages 321–340, Cham. Springer International Publishing.

[Veraart and Zdanowicz, 2015] Veraart, A. E. D. and Zdanowicz, H. (2015). Modelling and predicting photovoltaic power generation in the EEX market. *SSRN Electronic Journal:* https://ssrn.com/abstract=2691906.

[Virtanen et al., 2019] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . . (2019). SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. *arXiv e-prints*, page arXiv:1907.10121.

[Wind Europe, 2019] Wind Europe (2019). A new identity. https://windeurope.org/about-us/new-identity/. Accessed: 30.11.2019.

[Witherspoon and Fadrhonc, 2019] Witherspoon, S. and Fadrhonc, W. (2019). Machine learning can boost the value of wind energy. https://blog.google/technology/ai/machine-learning-can-boost-value-wind-energy/. Accessed: 01.12.2019.

[Zeileis and Hothorn, 2002] Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3):7–10.

[Zhu et al., 1997] Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 23(4):550–560.

[Zhu and Genton, 2012] Zhu, X. and Genton, M. G. (2012). Short-term wind speed forecasting for power system operations. *International Statistical Review*, 80(1):2–23.