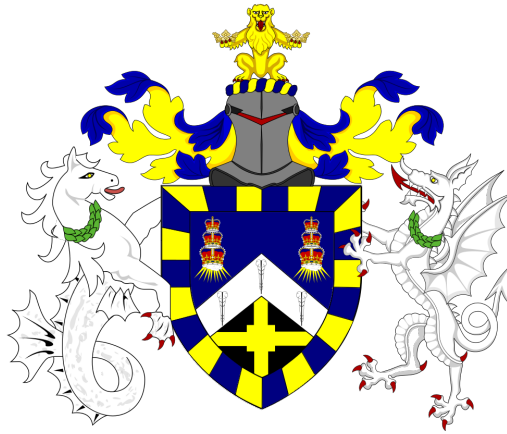


Time Series Analysis

An analysis of Nordic electricity spot prices and
market data

Antonio Emilio Invernizzi, ID 150633008

Advisor: Dr. Wolfram Just



A thesis presented for the degree of
Master in Mathematical Sciences

School of Mathematical Sciences
and School of Economics and Finance
Queen Mary University of London

Declaration of original work

This declaration is made on September 7, 2016.

Student's Declaration: I Antonio Emilio Invernizzi hereby declare that the work in this thesis is my original work. I have not copied from any other students' work, work of mine submitted elsewhere, or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Referenced text has been flagged by:

1. Using italics, **and**
2. using quotation marks "...", **and**
3. explicitly mentioning the source in the text.

Acknowledgements

I would like to express my gratitude to my supervisor Dr. Wolfram Just, who provided me the data to analyse and helped me throughout the process of writing this dissertation.

Abstract

The spot prices of electricity and the corresponding logarithmic returns provided by the Nord Pool exchange are analysed. Tests for mean, normality and stationarity are conducted. Some economic aspects, such as the dynamics of demand and supply of electricity in the Nordic European countries, and mean reversion are discussed by means of trends in the autocorrelation function. Attention is also devoted to the analytic rate of the decay of the autocorrelation function. Finally, the “on-off intermittency” behaviour of the logarithmic returns is studied and specific focus is paid to the probability distribution of laminar phases.

Contents

1	Introduction	5
2	Data of Nord Pool	8
2.1	Empirical findings and contributions	8
2.2	Description of the data	10
3	Preliminary Computations	15
3.1	Review of statistical concepts	15
3.2	Tests for the mean and normality	18
4	Analysis of Autocorrelation Function	22
4.1	Weakly stationarity	22
4.2	Autocorrelation of logarithmic returns	25
5	On-off Intermittency	32
6	Conclusions	38
A	Programming Code in R	40
B	Additional Tables	49
C	Additional Explanatory Figures	51
C.1	Daily, weekly and yearly trend	51
C.2	Decays of autocorrelation functions	52
C.3	On-off intermittency for different thresholds	54

Chapter 1

Introduction

Electricity is extremely different to traditional financial instruments and other commodities such as oil, grains, and metals. The main difference is that the former cannot be stored. As a consequence, electricity cannot simply be manufactured, transported and delivered. [14] Therefore, it has to be consumed immediately after it is produced and, in this sense, it can be considered more as a service than a good. This property makes the electricity market highly volatile and it is not surprising to observe a large number of extreme price movements. Another consequence of the non-storability of energy is a strong seasonality pattern in the data, which is discussed in further detail later. These problems also make it more difficult to price derivatives with electricity spot prices as the underlying. [1]

In electricity markets, a delicate balance between demand and supply must always be maintained in order to set spot prices in equilibrium. However, in practice, it is very difficult to steadily maintain such a balance. The supply stack is characterised as the ranking of all production units of a specific commodity in a given area. *“Electric power is usually generated from natural gas, coal, oil, nuclear fuel, falling water, geothermal steam, alternative resources such as cogeneration and from renewable resources such as wind power, solar energy, and biomass.”* [14]

The supply stack has a dynamical nature and changes over time since there are many factors that influence it. The demand for electricity, on the other hand, is largely weather dependent. It presents seasonal fluctuations, which are due to climate conditions and the need for electricity at a specific hour of the day. In Europe, the highest demand usually occurs in the winter due to excessive heating. In other geographical regions, like the southern states in the US, demand is at its highest in the summer since humidity and heat are the main reasons for

the excessive use of air-conditioning. [17]

In general, before the nineties, organizations employed long-term fixed price contracts for electricity. Those contracts, however, were over-the-counter, implying they were very inflexible and difficult to exit, once entered. Therefore, these organisations became increasingly exposed to price fluctuations. They had to hedge and trade electricity in a speculative way to tighten their exposure to price risk. In Norway, there was a need to get rid of the above-described issues. For this purpose, in 1991, the Norwegian parliament decided to deregulate the market for the trading of electrical energy. In this way, consumers could buy energy from other sources than their local provider. [11]

Norway was the first Nordic country which deregulated its power markets. The Energy Act of 1991 formed the basis for the deregulation of the power industry and as a consequence, other Nordic countries employed the same strategy. From this, a global trend started regarding the commoditization of electric energy. [14]

The deregulation enabled electricity to be traded in the same way as the other commodities, with risk management and trading as the main tools to run a successful business. [15]

In 1993, a Norwegian commodity exchange for electric power named Statnett Marked AS was established as an independent company. This provided the possibility to enter short- and long-term electricity contracts in addition to contracts for next-day or immediate physical delivery. [11]

The total volume during the first operating year was 18.4 TWh, at a value of NOK 1.55 billion.¹

Between 1992 and 1995 only Norway contributed to the market and in 1996 Norway and Sweden set up a common market for power. To expand its area of operations, Statnett Marked AS established a joint Norwegian-Swedish power exchange, which was renamed Nord Pool ASA. Svenska Kraftnt, an electricity transmission system operator in Sweden, became co-owner of Nord Pool ASA, which meant Statnett and Svenska Kraftnt each owned 50 percent of the Nordic power exchange and operated a common power exchange for the Nordic region.² In January 1998, it was estimated that of all trade in electric power in these two countries, around 25 percent was managed by Nord Pool. The majority of the electricity trade was still being organized as bilateral contracts between producers and consumers.

In general, the large industrial electricity consumers are more worried about unexpected increases in spot prices, whereas electricity producers are more concerned with large price drops

¹Please refer to “History” at <http://www.nordpoolspot.com/> for additional information.

²Please refer to the “Brief history” at <http://www.statnett.no/en/> for additional information.

than large price increases. The trading in futures and forward contracts would enable the above-mentioned actors to hedge their power portfolios and minimize the potential negative effect of an adverse fluctuation in electricity prices. [1]

For this purpose, electricity futures contracts began trading in Nord Pool, which was already the worlds most mature Scandinavian power market. It was then followed by the US NYMEX and the Australian/New Zealand markets. [16]

“Finland started a power exchange market of its own, EL-EX, in 1996 and joined Nord Pool in 1997”. On the 15th of June 1998, Finland became an independent price area in the Nord Pool Exchange. Consequently, Denmark also joined the exchange between 1999 and 2000. [4] At this time, after standard electricity options had already appeared in the NYMEX in 1996, Nord Pool introduced for the first time exchange traded Asian options on electricity futures. [16]

Over the years, Nord Pool expanded its activities and included even more participants, such as The Netherlands, the United Kingdom, Germany and the US. However, only the Norwegian and Swedish markets were fully integrated. [1]

Chapter 2

Data of Nord Pool

2.1 Empirical findings and contributions

After the deregulation of power markets, spot prices and returns of electricity have interested many researchers due to their contrast with traditional financial instruments. In particular, features such as price jumps, volatility, seasonality and mean reversion are recurrent in electricity markets.

The non-storability of energy is the main reason for extreme movements in prices. To measure and characterize the nature of such fluctuations, Byström [1] attempted to model electricity price changes with peaks-over-threshold and block-maxima methods. He applied these methods to the normal returns of spot prices from 1996 to 2000 using a combination of autoregressive model with a generalized autoregressive conditional heteroscedasticity (GARCH) model. He found that returns are highly not normally distributed with a large number of extreme observations. Weron et al. [15] applied regime switching models to the logarithm of the spot prices from 1997 to 2000, and confirmed the presence of jump behaviour.

Perelló et al. [9] used diffusion entropy analysis to logarithmic returns from 1992 to 2004. The method consists of converting the time series into a sort of probability density function and find a scaling exponent which may explain the nature of a spiky behaviour in the time series. They applied this method to a constructed stochastic process based on returns which are higher than a given threshold. By applying the above-mentioned technique, they demonstrated that the scaling parameters of the Nordic electricity spot prices are similar to those of the Dow-Jones and the US Dollar-Deutsche Mark futures. Simonsen [12] analysed the volatility of the logarithmic returns of Nord Pool from 1992 to 2004. He discovered that

the level of the daily volatility of the power market is 16%, which is very high compared to other traditional financial markets such as the stock and bond market. This is again because electricity cannot be stored in an efficient manner.

The mean reversion is another issue which interested several researchers. Such a property means that a market variable tends to revert to some long-term average level. In general, it is observable for instance in interest rates and commodities, particularly energy products. A reason for this is that once the prices of electricity increase, it will normally be consumed less and produced more, causing disequilibrium. This reduces the prices. On the contrary, as the price of electricity drops, it will probably be consumed more, but for the supply side, it will not be economically advantageous. This will cause the prices to increase. ([7], page 804)

Weron, Simonsen, and Wilman [17] applied the “Average Wavelet Coefficient” method to the spot prices of Nord Pool from 1992 to 2000, which highlighted that the time series is mean reverting. Very similar methods were used for the same set of data by Simonsen in the paper [11] with matching results. The mean reversion has also been found by Weron, Bierbrauer, and Truck [15] with regime switching models, for the years 1997 to 2000. This property seems to be present not only in Nord Pool but also internationally when considering electricity. Indeed, Weron, and Przybylowicz [14] discovered mean reversion for electricity prices of the Californian (CalPX) and Swiss (SWEP) exchanges using Hurst analysis. In the paper [16] by Weron, the autocorrelations of the lagged differences of the logarithmic returns of CalPX confirmed the mean reversion. Finally, Erzgräber et al. [4] analysed the hourly logarithmic returns of Nord Pool from 1992 to 2007 and attempted to determine whether the data presents long-range correlations using Hurst exponents. Without going into too much detail, such an analysis consists of computing the so-called R/S ratio and finding a specific power law which such a ratio should present. The ratio should behave as $R(\tau)/S(\tau) \sim \tau^H$, where τ is a time window of the series considered, and H is the Hurst exponent. The ideal case would be $H = 1/2$, meaning the series is a Brownian Motion, whereas if $H < 1/2$ or $H > 1/2$ the increments are anti-correlated or positively correlated, respectively. In this case, they did not find consistent and unequivocal results, since for different time ranges the Hurst exponent was different. For instance, exponents of 0.44, 0.33 and 0.23 were found for different time frames. These results showed that the time series is mainly anti-persistent. However, it was not possible to recognize a single Hurst exponent.

2.2 Description of the data

In this dissertation, data provided by the spot market operated by Nord Pool will be analysed with the statistical programming language R. For further details there is a section of the Appendix where the code applied to the data is shown.

As previously introduced, Nord Pool is an exchange market where participants trade power contracts for physical delivery the next day. Therefore, it is a short-term (one day) futures market. [1]

Elspot is the Nord Pool auction market which enables to trade electricity contracts. Players who want to buy/sell power on the Elspot market must send their purchase orders/sale offers to Nord Pool Spot the day before the power is delivered. The purchase orders and the sale offers are aggregated to demand and supply curve, respectively. The intersection of the two curves gives the market price for one specific hour. At noon, in the offices of Nord Pool in Oslo, computers calculate and publish the day-ahead prices. At the same time, they report to participants how much electricity they have bought or sold for each hour of the following day.¹

In addition to being an indicator of the current price for electricity, spot prices are also crucial for the pricing of derivatives. Forwards, futures, and options on electricity are traded in Nord Pool. When electricity spot prices serve as the underlying for these derivatives, they are usually collected over time; daily, weekly and monthly prices are computed from the hourly prices. [1]

There are 70,752 hourly spot prices, from which 70,751 hourly logarithmic returns can be calculated. From here, the 70,751 observations are denoted as n^* . Note that in one week and one year there are approximately 176 and 8,766 observations respectively. The analysis will focus on these returns, which are computed as

$$r_{\Delta}(t) = \ln(S(t)/S(t - \Delta)),$$

where $t \in \{1, 2, \dots, n^*\}$ is a specific hour inside the time frame between 01/01/1999 and 26/01/2007 and $S(t)$ denotes the spot price at hour t . From now on, the distance between hours will be $\Delta = 1$ and the analysis will focus on the logarithmic returns $r(t) = r_1(t)$.

In Figure 2.1 a time plot, which shows the data on the y -coordinate and the time index on

¹For additional information on this, please refer to pages 9 and 10 from the following booklet called “[The Nordic electricity exchange and the Nordic model for a liberalized electricity market](#)” from Nord Pool’s official website.

the x -coordinate, is shown. The electricity prices $S(t)$ are quoted in €/MWh (Megawatt hour). In general, it may be useful to determine whether there is a trend and seasonality to

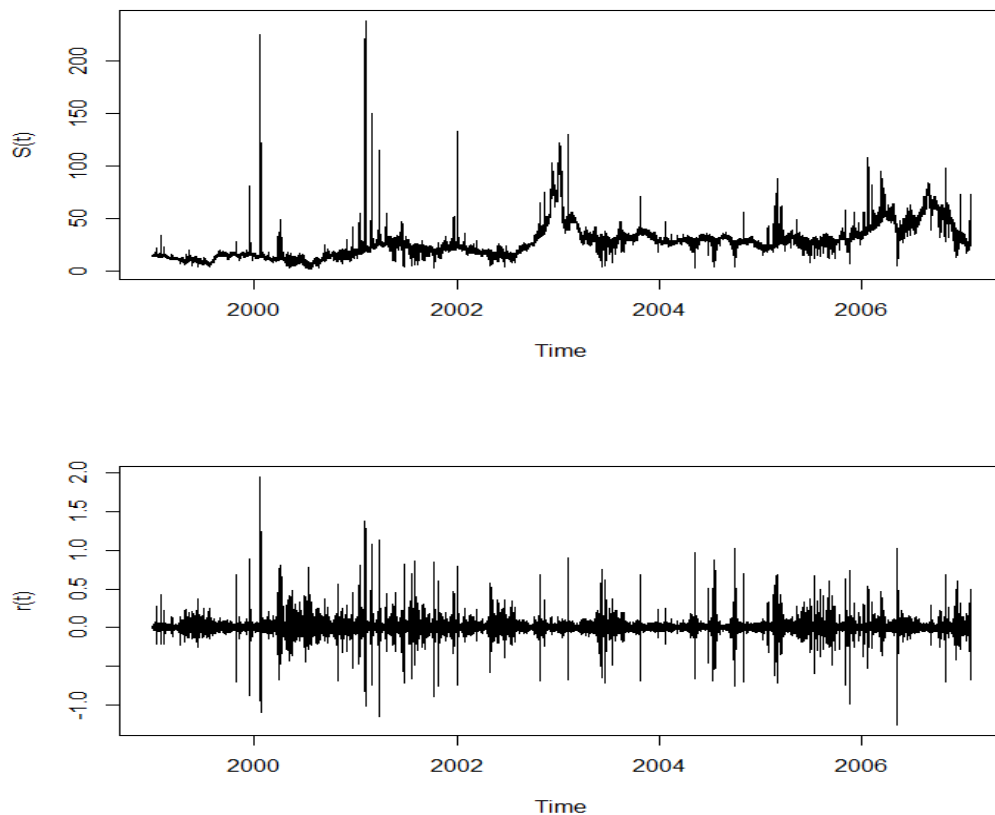


Figure 2.1: Top: Hourly spot electricity prices. Bottom: Hourly logarithmic returns from January 1999 to January 2007.

make predictions and create an adjusted set of data for simplicity purposes. A model which decomposes the spot prices in the trend and seasonality component is the following:

$$S(t) = m(t) + s(t) + \epsilon(t),$$

where at time $t \in \{1, 2, \dots, n^*\}$, $S(t)$ is the spot price, $m(t)$ is the trend, $s(t)$ is the seasonal effect and $\epsilon(t)$ is the remainder or residual series. This is also known as the classical decomposition model.

In order to perform a decomposition of the spot prices, first of all the trend $m(t)$ is estimated using a moving average over the years. This means that in the case of hourly data, with 8,766 observations per year, the yearly trend is estimated as:

$$\hat{m}(t) = \frac{r(t - 8766) + r(t - 8765) + \dots + r(t) + r(t + 1) + \dots + r(t + 8766)}{8766}.$$

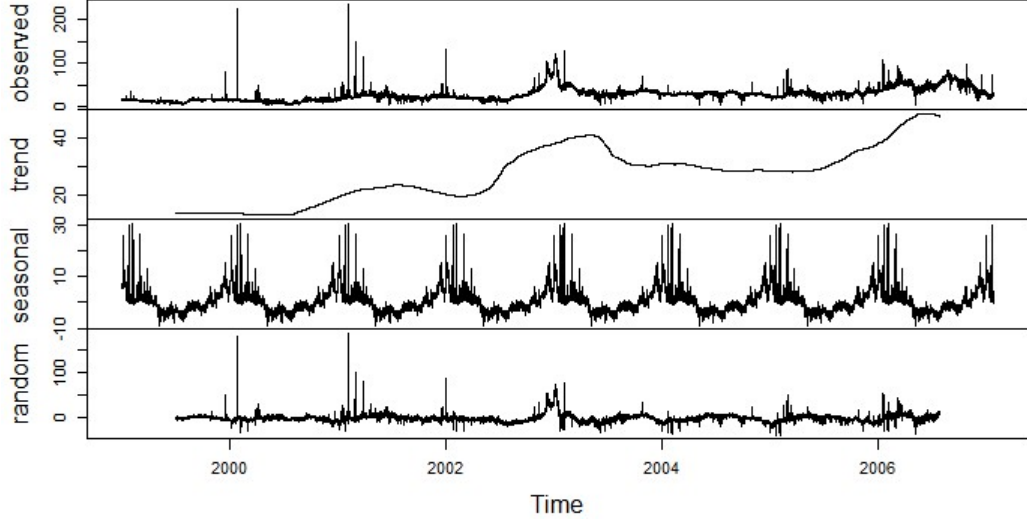


Figure 2.2: Decomposition of trend, seasonal and random components of electricity spot market prices.

In the Appendix, Figure C.1 shows how the daily, weekly and yearly trend calculated with moving averages for a better understanding and visualisation of the data.

To have a reliable estimation of the seasonal effect $\hat{s}(t)$, first one has to find $\hat{s}(t) = r(t) - \hat{m}(t)$. Then, by taking the average of the values for each season, it is possible to estimate a single component. For instance, for weekly data there may be n observations $\hat{s}(i), \hat{s}(168 + i), \dots, \hat{s}(168(n - 1) + i)$ for each week i , which can be combined to estimate the single weekly seasonal component as

$$\bar{s}(i) = \frac{\hat{s}(i) + \hat{s}(168 + i) + \dots + \hat{s}(168(n - 1) + i)}{n}.$$

Finally, the residual $\epsilon(t)$ can be obtained by subtracting the trend $\hat{m}(t)$ and seasonal components $\bar{s}(t)$ from the original spot prices $S(t)$. An alternative to finding the estimators described above is to fit a polynomial of order q , i.e.:

$$\hat{m}(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_q t^q$$

where $\alpha_0, \alpha_1, \dots, \alpha_q$ are estimated coefficients. If $q = 1$, this model reduces to a linear regression, which will be used later in the analysis.

In Figure 2.2 the decomposition of the logarithmic returns can be observed; the trend and seasonal components have been estimated using methods based on moving averages. ([2], pages 26-29)

The trend $m(t)$ is upward sloping throughout the years, with some falls at the beginning

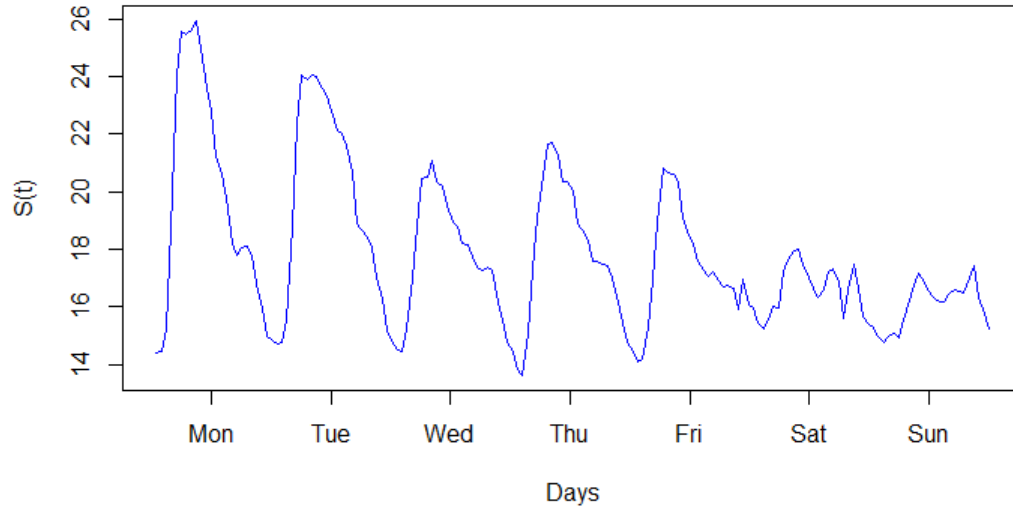


Figure 2.3: Electricity spot prices quoted in €/MWh for one week. The daily periodicity is observable.

of 2002 and mid-2003, and with the largest portions of growth observed between the end of 2002 and the beginning of 2003, and at the end of 2005. It appears to be difficult to understand the nature of these movements, but explanations for the mainly upward sloping trend can be found. One potential reason for this is that throughout the years more and more countries joined the Nord Pool exchange, causing a steady increase in the demand. Another reason may be inflation, which has a significant influence on all commodities' prices. Coming to seasonality, an evident seasonal component $s(t)$ can be observed. The peak of such a component is the beginning of the year, between autumn and winter, when days are shorter. In summer, the seasonal trend is lower. The most sensible economic reason for such a peak is that in winter, in Norway, people need more artificial light and heating, making the demand for electricity higher and causing the spot prices to increase. In the summer, when days are longer and the need for electricity is less pronounced, spot prices decrease accordingly. However, the winter peak of seasonality is not always the case across countries. For instance, in California, the highest consumption is in summer since there is a very high use of air conditioning and therefore much more electricity production is needed. [11]

The seasonality shown in Figure 2.2 is of a yearly nature. Looking at Figure 2.3, a daily rhythm is observable. During the day, the spot prices are higher due to increasing demand and high industrial activity. At night instead, there is not a great need for electricity and the spot

prices decrease accordingly. Electricity demand also has different dynamics throughout the week. During the weekdays from Monday to Friday the spot prices are higher than during the weekend. Clearly, there is more industrial activity during the weekdays and consequently this results in higher prices. Later, during the analysis of the autocorrelation function presented in Chapter 3, it will be possible to observe that there is also a seasonality of a weekly nature. Unexpected weather conditions can cause sudden and dramatic spikes in the prices. As soon as the weather phenomenon is over, the demand normally returns to its normal level. [17] This kind of behaviour, also present in the logarithmic returns, is an irregular variation of two states of the data. In one state, the returns behave quietly and there are no big surprises in the fluctuations. The other state is triggered by a burst in the returns, and such bursts can last for longer periods of time, for instance during adverse weather conditions. This kind of behaviour is also known as intermittency, which will be analysed in more detail later. However before analysing the complex features of the data, some fundamental values will be computed and commented in the next chapter.

Chapter 3

Preliminary Computations

In this chapter, the calculations of descriptive statistics are explained and some specific tests are performed, in order to come up with several considerations about the time series.

Before starting, a short review of the distributional properties of returns is exposed. The material for the next section is primarily taken from [\[13\]](#) pages 7-10.

3.1 Review of statistical concepts

Let X be a continuous random variable.

The expectation of X is formally defined as

$$\mu_X = E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

where $f_X(x)$ is the density of the random variable X . More generally the l th moment of a random variable is defined as:

$$E[X^l] = \int_{-\infty}^{+\infty} x^l f_X(x) dx.$$

The l th central moment of X is defined as:

$$E[(X - E[X])^l] = \int_{-\infty}^{+\infty} (x - \mu_X)^l f_X(x) dx.$$

The variance of a random variable X , which measures its variability, is defined as

$$\sigma_X^2 = \text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

The standard deviation of X is defined as the positive square root of the variance, σ_X .

The mean and the variance of a random variable uniquely characterise a normal distribution.

For other kinds of distributions, it may also be useful to calculate higher order moments.

The skewness measures the symmetry of X with respect to the mean and the kurtosis is used to measure the tail thickness of a random variable. They are respectively defined as:

$$S(X) = \mathbb{E} \left[\frac{(X - \mu_X)^3}{\sigma_X^3} \right], K(X) = \mathbb{E} \left[\frac{(X - \mu_X)^4}{\sigma_X^4} \right].$$

Since $K(X) = 3$ for a random variable X which is normally distributed, the quantity $K(Y) - 3$ is called excess kurtosis for a general random variable Y . Thus, the excess kurtosis of a normal random variable is zero. A distribution of a random variable with positive excess kurtosis is said to have fat tails, meaning that such distribution has a higher probability to present extreme results. Such a distribution is said to be leptokurtic. If a time series of returns presents a very high kurtosis, it is a signal that the normal distribution may be inappropriate. Note that both skewness and kurtosis are third and fourth normalised central moments.

In applications, estimators of the above defined statistics are used. In reality, it would be too difficult to construct a random variable which is continuous because it would have infinitely many values. Therefore, it is easier to sample random variables spaced between a certain time frame, which could be hourly (as in the case of the logarithmic returns), daily, monthly, etc. The following estimators for a random sample of n observations $X(1), X(2), \dots, X(n)$ of the variable X are used. These are estimators for the expectation, variance, skewness and kurtosis, respectively:

$$\begin{aligned} \tilde{\mu}_X &= \frac{1}{n} \sum_{t=1}^n X(t) \\ \tilde{\sigma}_X^2 &= \frac{1}{n-1} \sum_{t=1}^n (X(t) - \tilde{\mu}_X)^2 \\ \bar{S}(X) &= \frac{1}{(n-1)\tilde{\sigma}_X^3} \sum_{t=1}^n (X(t) - \tilde{\mu}_X)^3 \\ \bar{K}(X) &= \frac{1}{(n-1)\tilde{\sigma}_X^4} \sum_{t=1}^n (X(t) - \tilde{\mu}_X)^4. \end{aligned}$$

A traditional assumption made in financial studies is that the simple returns

$$R(t) = \frac{S(t) - S(t-1)}{S(t-1)}$$

for $t \in \{1, \dots, n\}$ are normally distributed, i.e. $R(t) \sim \mathcal{N}(\mu, \sigma^2)$, where the mean and the variance are fixed. This means that the returns have the following probability density function:

$$F(x) = P(R(t) \leq x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] dx,$$

with mean $E[R(t)] = \mu$ and variance $\text{Var}(R(t)) = \sigma^2$. However, the normal distribution encounters several problems and in reality, it is very difficult for the returns to behave as a normal distribution.

Another commonly used assumption is that the logarithmic returns are normally distributed. To explain briefly why it is so, the stochastic process $W = (W(t) : t \geq 0)$, called standard Brownian motion, is introduced.

The standard Brownian motion has the following properties:

- (i) $W(0) = 0$ almost surely.
- (ii) $W(t)$ is a continuous function of t .
- (iii) For $0 \leq s \leq t$ the increments $W(t) - W(s) \sim \mathcal{N}(0, t - s)$.
- (iv) For any finite set of times $0 = t_0 < t_1 < \dots < t_n$ the increments $W(t_i) - W(t_{i-1})$ are independent for all $i \in \{1, 2, \dots, n\}$.

The stochastic process which could better describe the movements of spot prices is the concept of geometric Brownian motion, where prices are defined by the following formula:

$$S(t) = xe^{Rt + \sigma W(t)}.$$

where x is a value greater than zero, $R > 0$ is the fixed interest rate, σ is the volatility and $W(t)$ is a standard Brownian motion.

With this notation, the ratio of the spot prices is given by:

$$S(t)/S(t-1) = e^{Rt + \sigma W(t) - R(t-1) - \sigma W(t-1)} = e^{R + \sigma(W(t) - W(t-1))}.$$

Applying the natural logarithm, it is simplified to

$$\ln(S(t)/S(t-1)) = R + \sigma(W(t) - W(t-1)),$$

with mean and variance:

$$\begin{aligned} E[\ln(S(t)/S(t-1))] &= R \\ \text{Var}(\ln(S(t)/S(t-1))) &= \sigma^2, \end{aligned}$$

since $W(t) - W(t-1) \sim \mathcal{N}(0, 1)$.

From this, it follows that $\ln(S(t)/S(t-1)) \sim \mathcal{N}(R, \sigma^2)$.

As will be seen in the next section, unfortunately with the data in question it is not possible to make the assumption that the logarithmic returns behave as a geometric Brownian motion.

3.2 Tests for the mean and normality

In this section, two questions have to be answered: Is the mean significant?

Is the time series normally distributed?

For the logarithmic returns of Nord Pool, the random process r is defined as

$r = (r(t) : t \in \{1, 2, \dots, n^*\})$. The results are shown in the following table:

Mean	Variance	Skewness	Kurtosis
6.69267e-06	0.003491878	1.930468	89.42235

Taking a quick look at the mean and kurtosis, one could make an approximate guess. Indeed, the mean is very close to zero and therefore one should expect that it is not significant. Also the huge kurtosis indicates the presence of fat tails, meaning that extreme events occur much more often than in a normal distribution.

To confirm these guesses, the following hypothesis testing for the mean is applied, namely:

- H_0 : $\tilde{\mu}_r = \mu_0$, “mean is equal to μ_0 ”.
- H_1 : $\tilde{\mu}_r \neq \mu_0$, “mean is different from μ_0 ”.

H_0 is called the null hypothesis, whereas H_1 is the alternative.

Using the test statistic

$$T(r) = \frac{\tilde{\mu}_r - \mu_0}{\sqrt{\frac{\tilde{\sigma}_r^2}{n^*}}} = \frac{\sqrt{n^*}(\tilde{\mu}_r - \mu_0)}{\tilde{\sigma}_r}$$

The H_0 hypothesis is rejected at the α significance level if $T(r) \in [z_{\alpha/2}, z_{1-\alpha/2}]$, where z_α is the α quantile of the normal distribution. The level α is chosen between the interval $(0, 1)$ and if $\alpha = 0.05$, this means that in 5 over 100 times, the test would wrongly reject the H_0 hypothesis.

Using the standard $\alpha = 0.05$, T should lie outside the interval $[-1.96, 1.96]$ to reject the null hypothesis.

For r , the test for the mean is done by setting $\mu_0 = 0$. The test statistic is

$$T(r) = \sqrt{70751} \frac{0.00000669267}{0.05909211} = 0.03012561.$$

Since $T(r) \in [z_{\alpha/2}, z_{1-\alpha/2}]$, it follows that the mean is not significant and H_0 is true. From this, the conclusion is

$$\tilde{\mu}_r = 0.$$

Now the test for the normality of the logarithmic returns is conducted. Looking at Figure 3.1,

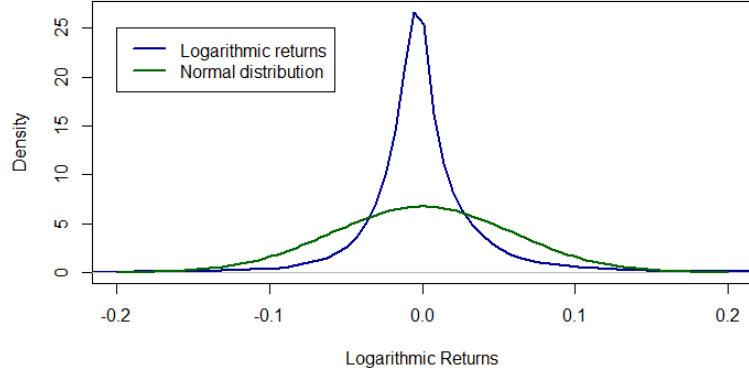


Figure 3.1: Comparison of the distribution of the logarithmic returns (black) with the normal distribution (blue).

one can definitely observe that the distribution of the logarithmic returns differs a lot from the normal distribution. However, it is very difficult from the figure to tell if the distribution is symmetric and if it has heavy tails. In order to do this, two other tests, namely the tests for skewness and kurtosis, are applied. The test for skewness is as follows:

- $H_0: \bar{S}(r) = 0$, “skewness is equal to 0”.
- $H_1: \bar{S}(r) \neq 0$, “skewness is different from 0”.

For this purpose, the following test statistic is used:

$$T(r) = \frac{\bar{S}(r)}{\sqrt{6/n^*}}.$$

The decision rule is the same as for the mean. The null hypothesis is rejected if $T(r) \notin [-1.96, 1.96]$. With the data in question, the test statistic for the skewness is

$$T(r) = \frac{1.930468}{\sqrt{6/70751}} = 209.6299$$

which obviously does not lie inside $[-1.96, 1.96]$. Therefore, the logarithmic returns do not have a symmetric distribution on the contrary to a normal one. From Figure 3.1, one might have told that the distribution of the returns is symmetric, but the test drastically rejects this hypothesis.

The test for the kurtosis is:

- $H_0: \bar{K}(r) - 3 = 0$, “excess kurtosis is equal to 0”

- H_1 : $\bar{K}(r) - 3 \neq 0$, “excess kurtosis is different from 0”.

In this case, the following test statistic is used:

$$T(r) = \frac{\bar{K}(r) - 3}{\sqrt{24/n^*}}.$$

These test statistics are asymptotically normal random variables. Again, the decision rule is the same as before. Calculating the test statistic from the logarithmic returns, the result is

$$T(r) = \frac{89.42235 - 3}{\sqrt{24/70751}} = 4692.311$$

which differs dramatically from zero. This is an indication of the presence of fat tails. Again, from Figure 3.1, the heavy tails cannot be captured. It seems that the normal distribution has heavier tails. This is because a data set of 70,751 points is very large and it is not possible from an image to observe this detail. One reason for the fat tails is that the electricity market is a volatile market, where big price changes occur relatively often, and sometimes these may be very extreme. Nevertheless, these extreme changes may occasionally occur in the data, and for the rest of the time, the returns behave “quietly”. [1]

The extreme price changes may also be caused by the impossibility to store electricity. It is not possible to have a perfect balance between demand and supply, since as soon as electricity is produced, it should be immediately consumed.

Another issue is that in the image only the $(-0.2, 0.2)$ range of the x -coordinate is shown. Some of the above-described extreme returns are not included in this range.

Jarque and Bera combine the above two tests in just one test:

- H_0 : “the time series is normally distributed”.
- H_1 : “the time series is not normally distributed”.

To apply this test, the following statistic is used:

$$JB(r) = \frac{\bar{S}^2(r)}{6/n^*} + \frac{(\bar{K}(r) - 3)^2}{24/n^*}.$$

The Jarque Bera statistic is asymptotically distributed as a chi-squared random variable with 2 degrees of freedom. The hypothesis H_0 is rejected if the p -value of $JB(r)$ is less than the significance level $\alpha = 0.05$. In other words, the p -value of a statistic is an estimation of the probability that a more extreme result than the result observed has occurred by chance, supposing that the null hypothesis is true. [3]

To explain this in more detail, suppose that an arbitrary random variable X has the same

distribution as the Jarque-Bera test statistic. The p -value is the probability that $|X|$ is bigger than $JB(r)$. Therefore, if a very small p -value is observed, it is very unlikely to get more extreme values than the test statistic and the null hypothesis could be rejected.

For the logarithmic returns, the Jarque-Bera statistic is $JB(r) = 22062000$ and the p -value is very close to zero and obviously smaller than the significance level 0.05. As a consequence, the hypothesis H_0 has to be rejected. From the tests conducted above, one can see that in this case, the normal distribution may not correctly describe the set of logarithmic returns.

Chapter 4

Analysis of Autocorrelation Function

Correlations between the variable of interest and its past values is one of the main issues of linear time series analysis. These correlations are known as serial correlations or autocorrelations. They are of fundamental importance for studying a stationary time series. [13]

In this chapter, tests regarding the stationarity of the time series in question are conducted and the autocorrelation function of the logarithmic returns is analysed. Before this, there is a quick introduction to covariance and correlation from probability theory.

The covariance of two random variables X and Y is defined as

$$\text{Cov}(X, Y) = \text{E}[(X - \mu_X)(Y - \mu_Y)] = \text{E}[XY] - \mu_X\mu_Y,$$

where $\text{E}[X] = \mu_X$ and $\text{E}[Y] = \mu_Y$.

The correlation coefficient of two random variables X and Y is defined as

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

4.1 Weakly stationarity

The covariance of the logarithmic returns is denoted as

$\gamma(k) = \text{Cov}(r(t), r(t - k))$ and is called the lag- k autocovariance of $r(t)$.

It has two important properties:

(i) $\gamma(0) = \text{Var}(r(t))$

(ii) $\gamma(-k) = \gamma(k)$.

A main concept for studying autocorrelations is the weakly stationarity. The returns $r(t)$ are said to be weakly stationary if:

(i) $E[r(t)] = \mu_r$ for all $t \in \{1, \dots, n\}$, the mean does not depend on time.

(ii) $\text{Cov}(r(t), r(t-k)) = \gamma(k)$, the autocovariance depends only on the k -lag.

In this dissertation, the term stationarity is linked to weak stationarity.

A descriptive interpretation of stationarity is that a stationary stochastic process is in a “stochastic equilibrium”, meaning that the subsections of a sample path have similar features, the results obtained across the subsections are consistent and the statistical fluctuations are purely random. [8], [13]

The empirical or sample autocovariance of the time series data $r = (r(t) : t \in \{1, 2, \dots, n^*\})$ is defined as

$$\tilde{\gamma}(k) = \frac{1}{n^*} \sum_{t=1}^{n^*-k} (r(t+k) - \tilde{\mu}_r)(r(t) - \tilde{\mu}_r), \quad 0 \leq k < n^*.$$

The empirical or sample autocorrelation function (ACF) of the time series r is defined as

$$\tilde{\rho}(k) = \frac{\tilde{\gamma}(k)}{\tilde{\gamma}(0)}, \quad 0 \leq k \leq n^*.$$

The question to be answered is: Is the time series of logarithmic returns stationary? As mentioned previously, the stationarity is characterised by consistent results within the subsections of a sample path. A straightforward example for this would be that the mean of the data is equal to zero, not only for the whole period, as already tested in Chapter 2, but also for every year. Looking at Figure 2.1, it appears that the mean for every year is zero within statistical fluctuations. However, from 2000 to 2002 there are several extreme movements. It is therefore better to conduct hypothesis testing again. The values for every year are shown in Table B.1 in the Appendix. The testing procedure is the same as for the mean. Again, one can see that here all these test statistics also lie inside the interval $(-1.96, 1.96)$, confirming that for every year, the mean is not significant.

So, from this, it can be supposed that the series is stationary. However, more evidence is needed to confirm the stationarity.

To do this, another test called the “differences sign test” is introduced. The following has to be tested:

- H_0 : “No trend in the data”.

- H_1 : “There is a trend in the data”.

First, the set S is defined as

$$\begin{aligned} S &:= \{i : r(i) > r(i-1), i = 2, \dots, n\} \\ &= \{i : r(i) - r(i-1) > 0, i = 2, \dots, n\} \end{aligned}$$

for a series of independent, identical distributed random variables, the cardinality of S is binomial distributed, namely $|S| \sim \mathcal{B}(1/2, n-1)$ with

$$\begin{aligned} \mu_S &= E[|S|] = \frac{1}{2}(n-1) \\ \sigma_S^2 &= \text{Var}(|S|) = \frac{1}{12}(n+1) \end{aligned}$$

and most important is that

$$T(S) = \frac{|S| - \mu_S}{\sigma_S}$$

converges in distribution to a standard normal distribution $\mathcal{N}(0, 1)$ as n goes to infinity. Again, taking the significance level α as 0.05, H_0 has to be rejected if $T(S) > z_{1-\alpha/2} \approx 1.96$. Such a testing procedure is also described in [8] page 12.

Below is a direct application of the test to the data in question. However, it would be difficult to apply it to the complete set of data with $n^* = 70,751$ observations, because the data presents evident seasonality and this would greatly affect the reliability of this test. Indeed, the seasonality effect is that in the data there are frequent switches of direction of the logarithmic returns, making the cardinality of S too large or too small. In this case,

$$\begin{aligned} \mu_S &= \frac{1}{2}(70751 - 1) = 35375 \\ \sigma_S^2 &= \frac{1}{12}(70751 + 1) = 5896 \\ \sigma_S &= \sqrt{5896} \approx 76.78. \end{aligned}$$

Performing the calculation of $|S|$ with n^* observations, the cardinality of S is 37,177, resulting in more upswings in the returns than decreases. Therefore,

$$T(S) = \frac{|S| - \mu_S}{\sigma_S} = \frac{37177 - 35375}{\sqrt{5896}} \approx 23.468.$$

The hypothesis H_0 is rejected since 23.468 is far larger than 1.96.

To partly get rid of the seasonality problem, the same test for every year from 1999 to 2006 is applied and only the daily returns of a specific hour are considered. In this case there are

365 observations and

$$\begin{aligned}\mu_S &= \frac{1}{2}(365 - 1) = 182 \\ \sigma_S^2 &= \frac{1}{12}(365 + 1) = 30.5 \\ \sigma_S &= \sqrt{30.5} = \frac{\sqrt{122}}{2} \approx 5.522.\end{aligned}$$

The cardinalities of S and $T(S)$ for the years from 1999 to 2006 are shown in Table B.3 from the Appendix. In the year 2001, the value 1.9917864 is slightly greater than 1.96, so in this case it might be difficult to determine whether there is a trend. However, it is still close to 1.96 and this is why in this case the H_0 hypothesis will not be rejected. Looking at Figure 2.1, it can be observed that the most extreme price changes occur in 2001. For all the other years, the H_0 hypothesis is definitely not rejected. From this, the fact that the data is stationary is accepted.

4.2 Autocorrelation of logarithmic returns

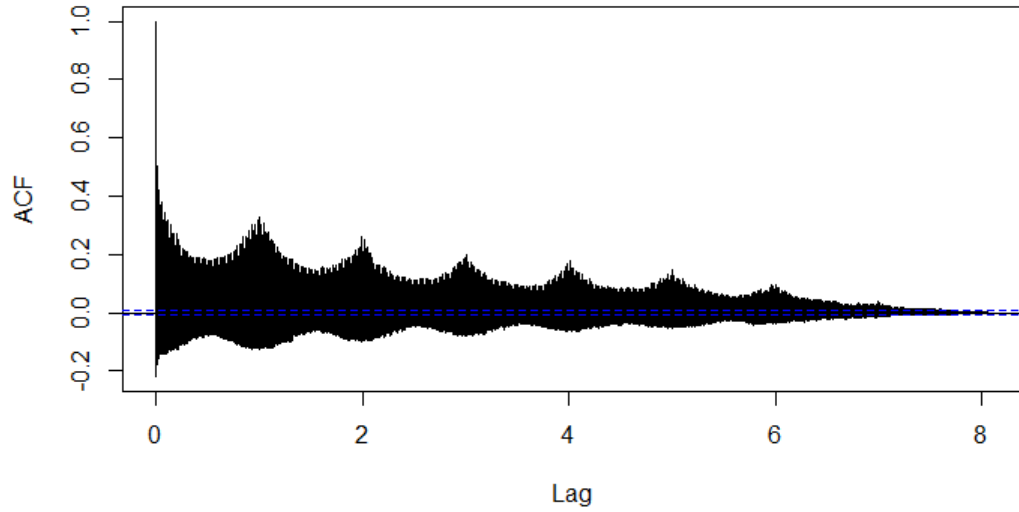


Figure 4.1: Autocorrelation function for n^* lags.

In Figure 4.1, the autocorrelation function with n^* lags is shown. The averages

$\tilde{\rho} = (\tilde{\rho}(k) : k \in \{0, 1, \dots, n^*\})$ have been computed according to the following equation, i.e:

$$\tilde{\rho}(k) = \frac{1}{n^* \tilde{\gamma}(0)} \sum_{t=1}^{n^*-k} (r(t+k) - \tilde{\mu}_r)(r(t) - \tilde{\mu}_r), \quad 0 \leq k < n^*.$$

where $\tilde{\gamma}(0)$ is the variance of r . In the x -coordinate, the lags are shown in terms of years for the sake of simplicity. Each year, which in the figure is represented as one unit, comprises approximately 8,766 lags. A clear yearly rhythm is observable. Under the condition that the picture displays reliable results, it can be argued that the data shows long-range correlations. However, from the figure, it is very difficult to capture another kind of correlations, for instance within days and weeks. To make such correlations observable, a period of respectively 24, 168, 336 and 672 lags is analysed.

Suppose that, for one week, the price of electricity is in equilibrium, which means that the supply perfectly meets the needs of the demand. In such a market, extreme weather conditions which would seriously affect production and consumption patterns, are excluded. The expectation of the daily trend of electricity spot prices in Northern Europe, during a weekday, would be the following. From approximately 11 p.m. to 6 a.m., there is hardly any demand for electricity since people usually sleep during these hours and industrial activity is low. [17] In the morning, from 6 a.m. to 10 a.m., a considerable increase in the prices would occur since the demand rises in the morning when people wake up and start working. Then from approximately 9-10 a.m. to 6 p.m., one would predict that electricity prices follow a stable trend without too many sudden fluctuations since industrial productivity mainly remains constant for those hours. After this, from 6 p.m. to 11 p.m. there should be a steady decrease in prices. Those phases where the prices increase or decrease more than usual can be named transition phases. During the day, there are two of these phases, one in the morning and one in the evening, and they last approximately 4-5 hours each. This means that electricity prices are in a transition phase almost 8 hours per day. The more stable phases are respectively from 10 a.m. to 5 p.m. and from 11 p.m. to 6 a.m.

When prices are in equilibrium, the above-described cycle should repeat itself throughout the weekdays. As already mentioned, this kind of trend may differ in magnitude at the weekend due to different patterns of demand and supply.

In a power market, one would also expect to have recurrent events every 24 hours, since demand dynamics at a specific hour of one day are very similar to the dynamics of one day later, in the same hour. Also, a weekly trend would not be surprising since the intensity of work on a specific day from Monday to Sunday repeats itself throughout the weeks. For instance, on Monday it is expected to have more industrial activity than on Friday or Saturday. How is it possible to empirically observe the above-described scenarios? One answer could be related to the autocorrelation function.

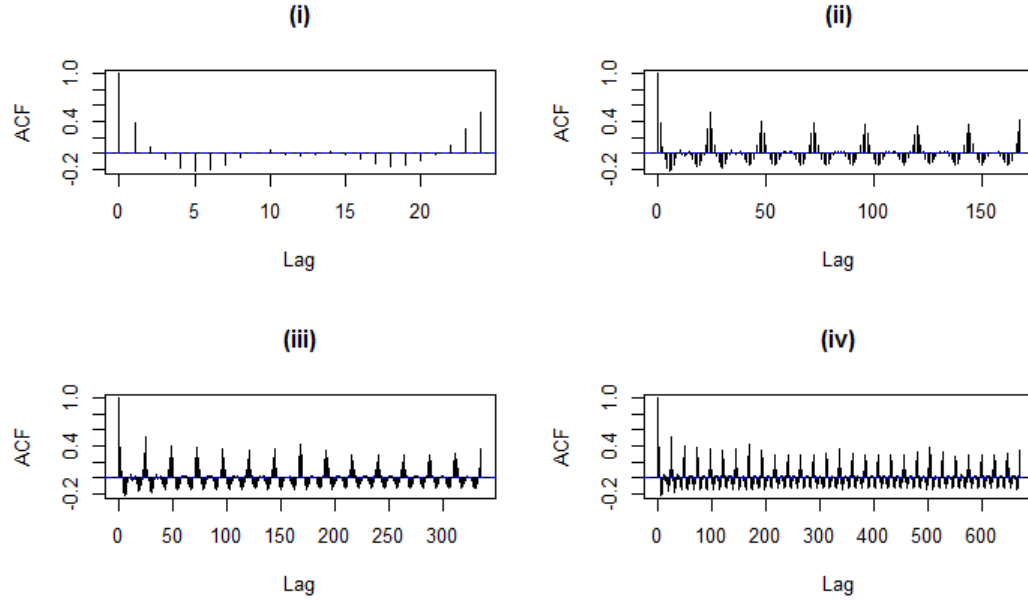


Figure 4.2: Autocorrelation function of hourly logarithmic returns of Nord Pool electricity spot prices with (i) 24 lags (one day), (ii) 168 lags (one week), (iii) 336 lags (two weeks) and (iv) 672 lags (one month).

After the reasoning above, it is expected to have higher positive correlations every 24 hours (daily) and 168 hours (weekly). Figure 4.2 shows the autocorrelation function for several time lags. In (ii), the expected daily positive correlation is observable since there are 7 “humps” at every 24-hour distance, meaning that movements in the spot prices of one day at a specific hour are correlated with the movements of prices of the day before at the same hour. In (iii) and (iv), even the expected weekly positive correlation is observable. The movements of the positive daily correlations within one week are comparable to a parabola function. Indeed, the positive correlation decreases slightly after the first 4-5 days, before increasing again until the 7th day.

So far, it seems that the autocorrelation function confirms the intuitions made above. However, there are some surprises caused by unexpected movements in the autocorrelation function within one day. Looking again at Figure 4.2 (i), two well pronounced negative correlations are observable at time lags 6 and 18. Such negative correlations are somehow unanticipated. Every 6 hours, the returns behave mainly in an opposite way. The reason for this could be that in the electricity prices, both transition phases mentioned above could last for several hours per day, approximately 4 hours each. If the considered hour is in the range from 6 a.m. to 10 a.m, six hours later the returns would escape from the transition phase and enter a

more stable phase. Therefore, in the stable phase, smaller returns would be observable, and these could cause a negative correlation with the larger returns observed during the transition phase.

Instead, from 12 p.m. to 5 p.m. the returns would turn from the stable to the transition phase of the evening, from 6 p.m. to 11 p.m. Continuing with this reasoning for every hour of the day, it would be possible to determine that the only immune ranges from the change of phases would depart from 10 a.m, 11 a.m, 11 p.m. and 12 a.m. These are only 4 out of 24 possible ranges. The other 20 ranges clearly have a more intense weight in the computation of the autocorrelation function. The second well pronounced negative correlation is in the 18th lag, i.e. 12 hours after the first one. In this case, it is possible to make the same reasoning as before. The only difference is that if returns are in a stable phase, after 18 hours they will not enter in the first coming transition phase, but rather in the second. This is one possible explanation for the negative correlation in the 6th and 18th lags.

Interestingly at lag 12, the returns are uncorrelated. This makes absolute sense since the demand and supply dynamics during the day are completely different and independent from those during the night. The exception could come from the ranges 7 a.m.-7 p.m. and 8 a.m.-8 p.m. since they are in between day and night. However, these are only 2 over 24 ranges.

When analysing autocorrelations, it is also important to consider the property of mean reversion. Since the one-day-ahead dependencies in the autocorrelation last for almost all the data points in question, it is reasonable to consider the daily logarithmic returns differenced by a lag of 7, again in order to adjust for seasonality. For the whole period, from the beginning of 1999 to the beginning of 2007, the $n^* = 70,751$ hours correspond to approximately $\tilde{n} = 2,947$ days. To apply the differencing method explained above, one considers the modified returns $z(t^*) = r(t^* + 7) - r(t^*)$ for $t^* = 24t$, where $t \in \{1, 2, \dots, \tilde{n}\}$. The mean reversion can be detected once the covariance of differences of logarithmic returns over two disjoint periods is negative. ([5], page 22)

More precisely, this means that

$$\text{Cov}(r(s+7) - r(s), r(u+7) - r(u)) < 0 \text{ for all } t < t+7 \leq u < u+7.$$

In [16], it was discovered that mean reversion applies to the returns of the California Power Exchange (CalPX) after differentiation of 7 days. Therefore, since electricity markets behave in a similar manner across the world, it is expected that the logarithmic returns of the Nord Pool exchange will also show mean reversion. However, it has to be noted that in California the peak of electricity prices comes in summer, due to the strong use of air conditioning,

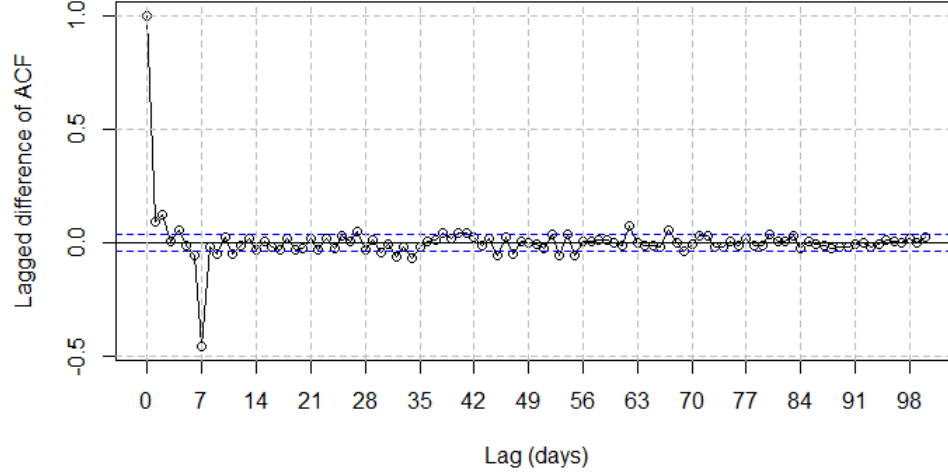


Figure 4.3: Lagged autocorrelation function for Nord Pool logarithmic returns after differentiation by 7 days.

while in Norway it is in winter, due to a greater need for heating. Figure 4.3 shows the autocorrelation function of z_t . The blue dashed lines show the 95% confidence interval for the autocorrelation function, i.e.: $\left[-1.96/\sqrt{\tilde{n}}, 1.96/\sqrt{\tilde{n}}\right] = [-0.03614784, 0.03614784]$. The interval is chosen as this since it can be shown that $\sqrt{\tilde{n}}\tilde{\rho}(k) \sim \mathcal{N}(0, 1)$ for a considerably large number of observations.¹ There are some small positive correlations in the first two lags and a negative correlation in the seventh lag. The remainder of the autocorrelation function is not significant. Therefore, also in the case of the Nord Pool logarithmic returns, the mean reversion is observable since the autocorrelation at the 7th lag is negative and not significant after it.

Another interesting feature of the autocorrelation function is the rate of the decay over several years. Normally, when lags k are considerably large, then $\gamma(k)$ tends to decay and approach to zero. Also, in the case of the autocorrelation function of logarithmic returns, this decay is observable in Figure 4.4. However, it decays rather slowly over the years. To understand the analytic rate of decay of the autocorrelation function, only every 24 hours are considered to suppress the daily periodicity. The simplest way to recognize the nature of the decay is to check whether the autocorrelation is linear on a semi-logarithmic scale, meaning it is exponential, or linear in a double-logarithmic scale, and it is a power-law. In Appendix C.2,

¹For more details, please refer to [13] page 27.

the decay for the daily autocorrelation for every year is observable, both on the linear and double logarithmic scale. There is a very similar decay across the autocorrelation functions. When looking at the plots, on the double logarithmic scale the trend of the decay seems to be linear at the beginning, and then it plummets at the end. This may suggest there is a cutoff k^* in the lags, where for all $k < k^*$, the decay follows a power law, and for all $k \geq k^*$ the decay is rather exponential. In the case of the daily autocorrelation function with n^* observations, k^* is found to be approximately 1,700. Here, lags smaller than the cutoff $k^* \approx 1700$ are considered. In order to explicitly compute the rate of the decay, one needs to determine the value of r from the following equation:

$$\tilde{\rho}(k) = \lambda k^r.$$

Applying the natural logarithm to the above formula, one obtains:

$$\log(\tilde{\rho}(k)) = \log(\lambda k^r) = \log(\lambda) + r \log(k).$$

Note that this formula has the same form as a linear regression

$$y = \alpha_0 + \alpha_1 x,$$

where the dependent variable y is $\log(\tilde{\rho}(k))$, the independent variable x is $\log(k)$ and the coefficients are $\alpha_0 = \log(\lambda)$ and $\alpha_1 = r$. To find α_1 , a linear regression of the above form is performed. In Figure 4.4, the decay of the autocorrelation with n^* observations on a double logarithmic scale is observable. Applying the above-described linear regression, it turns out that the decay of the autocorrelation function α_1 is approximately a power law of $-0.30794553 \approx -0.3$, and the intercept α_0 is $-0.09331027 \approx -0.09$. In order to find λ , one just needs to calculate $e^{\alpha_0} \approx 0.91$. Summing up, the absolute value of the autocorrelation function with n^* observations behaves approximately as:

$$\tilde{\rho}(k) \sim 0.91k^{-3/10} \text{ for } k < k^*.$$

Instead, if one considers the decay just for the autocorrelation of one year, different results would be observable. It turns out that the decay of the autocorrelation function over several years is also a power law. However, the rate of such decay is different across the years and ranges approximately from $-3/5$ to $-2/5$. A computation of every rate of decay has been computed and can be found in Table B.2 in the Appendix.

In section C.2 in the Appendix, one can also find the plots of decay of the autocorrelations are shown with the respective power law for better understanding.

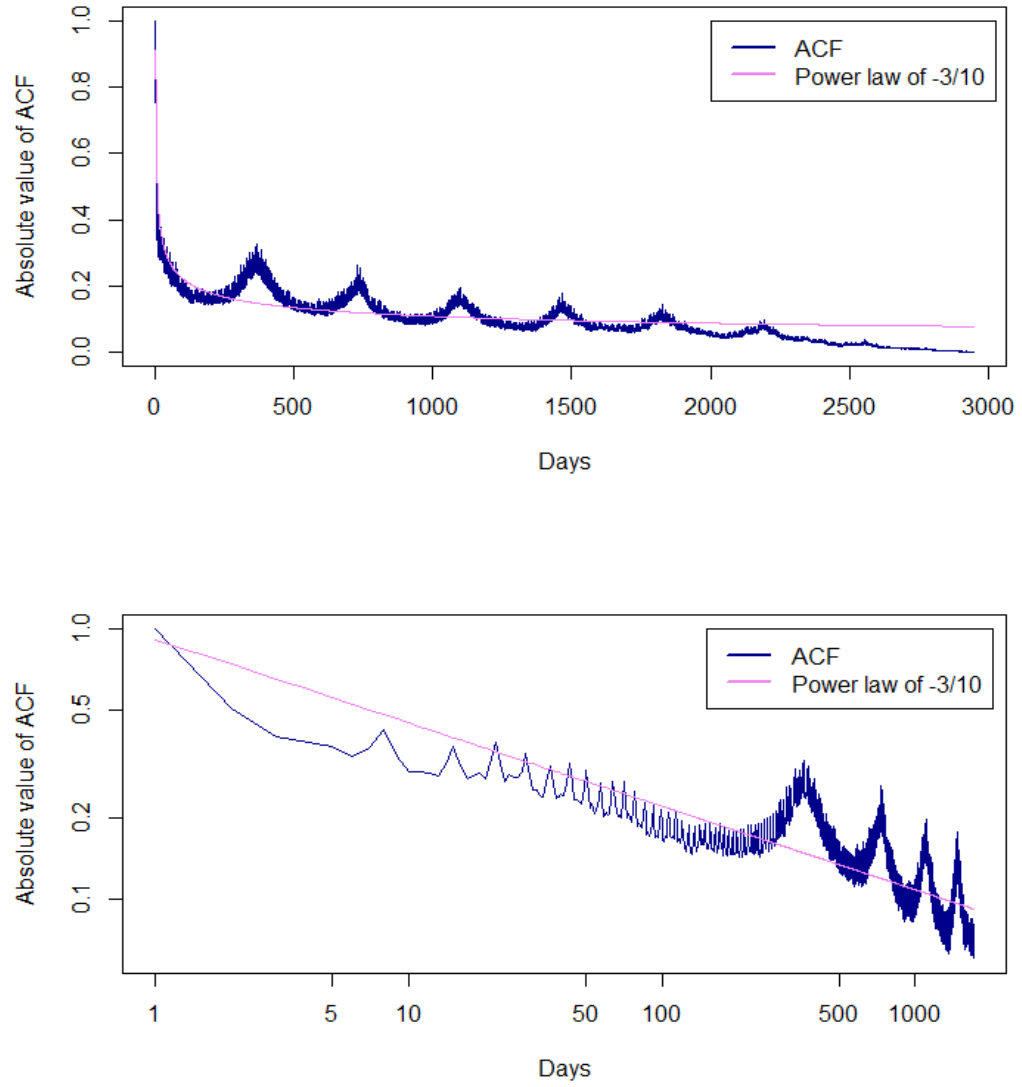


Figure 4.4: Plot of the absolute value of the daily autocorrelation function for the whole period in both the normal and double logarithmic scales.

Chapter 5

On-off Intermittency

In this chapter, a specific kind of intermittent behaviour of a stochastic process, known as “on-off intermittency”, is explained. The paper [6] deals with the characterisation of on-off intermittency and explains in detail the underlying theory.

Such behaviour has two different states. When the time series is in the so-called “off-state”, it means that it is almost constant within statistical fluctuations and quite close to its mean. To characterise the “off-state”, one needs to find suitable thresholds for the left and right endpoint of an arbitrarily chosen interval. This interval should be centred around the mean of the series. After this, one needs to check whether the data lies within this interval and count how many consecutive points are in this state to characterize the lengths of the respective “off-states”. This state is nearly constant and can remain so for very long periods of time. Such events are also called laminar phases of a time series. The “on” state is a burst, departing quickly from, and returning quickly to, the off state.

The main aim is to find the distribution of laminar phases. But how can such laminar phases be characterised? In general, a laminar phase is defined as the number of consecutive observations which are in the off-state. For this purpose, one has to choose an arbitrary threshold $\epsilon > 0$ for a time series $Y = (Y(t) : t \in \{1, 2, \dots, n\})$ with mean μ_Y , and check how many consecutive observations lie within the interval $[\mu_Y - \epsilon, \mu_Y + \epsilon]$. Therefore, the laminar phases are series of consecutive numbers which are smaller than a chosen value. As soon as one value is larger than the threshold, the on-state starts and one should wait for the start of the next laminar phase. The laminar phases of the series Y are defined here as L_t , where $t \in \{1, 2, \dots, n\}$ is the starting point of the off-state. For example, consider the following series $\tilde{Y} = (1/10, 3/10, 7/10, 1/2, 2/5)$. Then the mean is $\mu_{\tilde{Y}} = 2/5$. If ϵ is chosen as

$1/5$, the interval reduces to $[1/5, 3/5]$. In this case only two laminar phases would be present, namely $L_2 = \{3/10\}$ and $L_4 = \{1/2, 2/5\}$ with respective lengths $|L_2| = 1$ and $|L_4| = 2$. Of particular interest is the probability that a laminar phase has precisely length τ . This probability is denoted as $\phi(\tau)$.

For many cases studied, it can be shown¹ that the asymptotic distribution of this probability fits well with a power-law with exponent $-3/2$. The meaning of this is

$$\phi(\tau) \sim \lambda \tau^{-3/2}$$

where $\lambda > 0$ is a constant.

Now, the above-described theory will be applied to the data in question. Looking back at Figure 2.2, it can be observed that the trend is almost linear from the middle of 2003 to the end of 2005. In the years from 2000 to 2002 instead, there are many extreme movements in the prices, as can be seen in Figure 2.1. Therefore, this is the period with the highest likelihood to present longer laminar phases. The logarithmic returns of the electricity prices of Nord Pool are particularly well suited to an on-off intermittency analysis since there is an abundance of extremely large fluctuations in the prices. Indeed, once the prices reach a peak, it could last just a few hours for them to return to the usual level. One of the most extreme jumps occurred on the 5th of February 2001, when the spot price reached an overall high of 1951.76 NOK/MWh. After only three hours, the spot price reduced to 200 NOK/MWh, which at that time was the normal level for the spot prices. In the case of Nord Pool, which includes Scandinavian countries, the period of major consumption is winter and this is the season where dramatic price variations happen. During low consumption periods, there are typically enough sources to produce energy and the supply can easily meet the demand. However, when there is a high need for electricity, “*only a minor increase in the demand can have dramatic consequences for the electricity prices*”. As soon as the demand is higher than expected, it is more costly to produce additional energy since the production units may not be sufficient. Clearly, this has a dramatic impact on the spot prices and implies the volatility is higher during winter seasons, and lower during summer seasons. [9]

To empirically observe that the largest fluctuations occur in winter, Figure 5.1 shows the frequencies of the 100 largest absolute values of the logarithmic returns observed for every month. One can observe that the more frequent extreme movements happen in January, February, and October. It is not surprising that the largest frequency is in February, since the event described above in February 2001 presented several consecutive large price movements,

¹Please refer to [6], section III for further details.

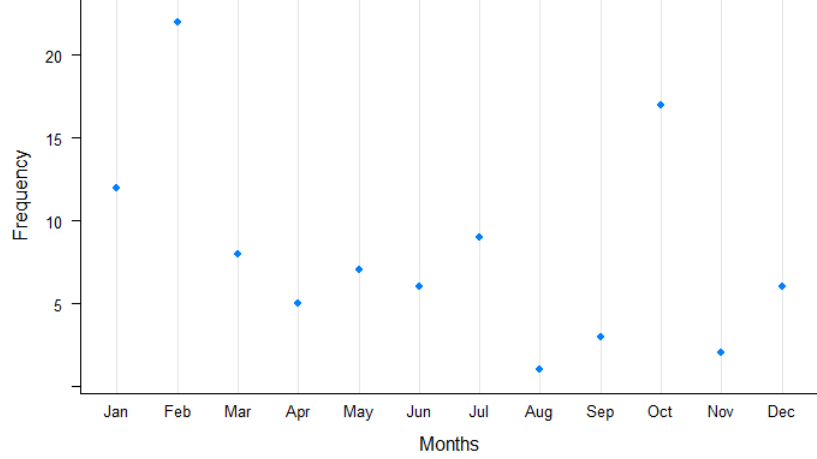


Figure 5.1: Frequencies of the 100 largest price fluctuations for every month.

making the frequency higher. A large number of observations in October is somehow unexpected, but is still sensible since winter is approaching and the season is changing. Therefore, the demand will increase more in October, compared to the previous months when the logarithmic returns are quite stable and the volatility is not particularly high, as shown in the figure.

Now it is time to show how the laminar phases behave in the logarithmic returns. As already mentioned, one should choose an arbitrary threshold $\epsilon > 0$, and check if the returns $r = (r(t) : t \in \{1, 2, \dots, n^*\})$ are within the interval $r(t) \in [\tilde{\mu}_r - \epsilon, \tilde{\mu}_r + \epsilon]$. From the test of the mean in Chapter 2, it holds $\tilde{\mu}_r = 0$. From this, the interval simplifies to $r(t) \in [-\epsilon, \epsilon]$. For the data in question, the threshold is chosen as $\epsilon \in \{0.05, 0.1, 0.2, 0.3\}$. The following modified logarithmic returns are defined:

$$r^*(t) = \begin{cases} |r(t)|, & \text{if } r(t) \in [-\epsilon, \epsilon] \\ 0, & \text{otherwise.} \end{cases}$$

In Appendix, the intermittent behaviour of r_t^* for $\epsilon \in \{0.05, 0.1, 0.2, 0.3\}$ is shown. The probability density function $\phi(\tau)$ is computed using the Kernel density estimation.²

After the Kernel density is computed, the zeros in the function are removed for simplicity purposes and an estimation $\tilde{\phi}(\tau)$ is computed. Finally, the main aim here is to figure out the analytic rate of decay of the probability density function for different ϵ s. In other words, the

²For additional information about the algorithm used by R to compute the density function, please refer to the following website: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/density.html>.

aim is to find r from the equation

$$\tilde{\phi}(\tau) = \lambda\tau^r,$$

The ideal case would be $r = -3/2$. Again, according to the procedure used for the autocorrelation function, applying the logarithm to the above formula, one obtains:

$$\log(\tilde{\phi}(\tau)) = \log(\lambda\tau^r) = \log(\lambda) + r \log(\tau),$$

which has the same form as

$$y = \alpha_0 + \alpha_1 x,$$

where the dependent variable y is $\log(\tilde{\phi}(\tau))$, the independent variable x is $\log(\tau)$ and the coefficients are $\alpha_0 = \log(\lambda)$ and $\alpha_1 = r$. To find α_1 , a linear regression of the above form is performed. In Figure 5.2, the decays of $\tilde{\phi}(\tau)$ for several ϵ s are depicted. For $\epsilon = 0.05$, it seems that the decay is rather fast. This makes sense since electricity markets are very volatile and therefore it is difficult to observe very long laminar phases because frequently returns are larger than 5%. For $\epsilon \in \{0.1, 0.2\}$, the decays look similar, but when the threshold is 0.2 it is clearly slower since it is more difficult for the returns to overcome the limit of 20%. This results in the observation of longer off-states. Along with the decay of laminar phases for $\epsilon = 0.2$, also a power law of $-3/2$ can be observed to compare how the decay differs from the desired result. Finally, with $\epsilon = 0.3$, the slowest decay is observed. The next table shows the coefficients α_0 and α_1 , or equivalently the intercepts and the analytic rates of decay for every ϵ .

ϵ	0.05	0.1	0.2	0.3
α_0	0.789407	0.0176894	-0.2107925	-1.060555
α_1	-1.993119	-1.6475308	-1.4279807	-1.213252

From these values, it may be possible to reach the following suppositions: For the threshold 0.05, there is a faster decay, and clearly far away from the desired $-3/2$ power law. This may suggest that $\epsilon = 0.05$ is not the most appropriate threshold to choose since too many bursts over 5% occur in the logarithmic returns for the reasons explained above. This would create too many laminar phases of length 1 and fewer laminar phases of larger length. For $\epsilon = 0.3$ instead, the decay is rather slow but also far from the desired result. Also, 30% is not an appropriate choice. There are still several returns larger than 30%, but these do not occur very often and cause the laminar phases to be too long. For thresholds larger than 0.3, it would not make sense to analyse “on-off intermittency”, since almost only off-states would be present and clearly the probability density function decays slower. The thresholds

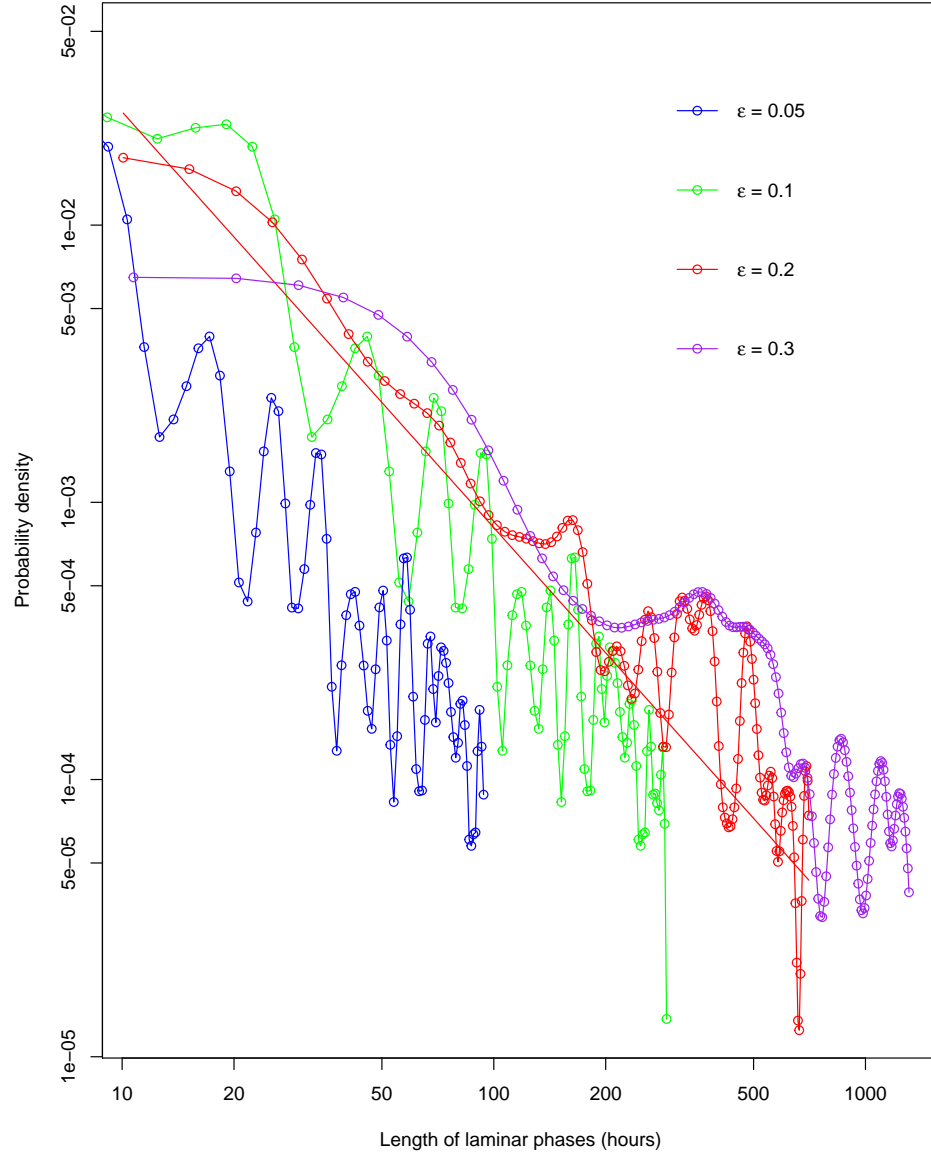


Figure 5.2: The probability density function $\tilde{\phi}(\tau)$ of laminar phases against the length of the time interval τ for different values of the threshold ϵ indicated in the legend, plotted in a double logarithmic scale.

$\epsilon \in \{0.1, 0.2\}$ seem to be the most reasonable choice, since the power laws of approximately -1.64 and -1.42 are very near to the desired result of -1.5. Since the desired power law lies in between the two observed results, this suggests that the best choice of threshold would be $\epsilon^* \in (0.1, 0.2)$. For such a threshold, the theory from [6] would perfectly describe the case of the logarithmic returns of electricity spot prices from Nord Pool.

Chapter 6

Conclusions

Some interesting aspects of the logarithmic returns of electricity spot prices quoted by Nord Pool were analysed.

First of all, the test for the mean delivered the result that logarithmic returns have an overall mean of zero for a considerable number of observations. Furthermore, it was also found that for every year the mean is zero. This fact with the “differences sign test” applied to the daily logarithmic returns of every year, confirmed that the time series under question is stationary. The returns were found to be highly non-normally distributed. This is due to their high positive skewness and huge kurtosis. The positive skewness indicates that in electricity markets returns tend to move upwards more than downwards. This property was also observed by the calculation of the trend, which mainly increases over the time frame considered. The fat tails indicate a significant presence of extreme price fluctuations in the electricity market. Such movements are not that surprising in such a market since energy cannot be stored efficiently. In addition, the dynamics of demand and supply are highly weather dependent. If an unexpected weather phenomenon occurs, often it would be very difficult to produce as much electricity as is needed. Therefore, dramatic jumps would happen, and it may also be that after a few hours the prices return to their normal levels. Clearly, this would result in very high consecutive returns for a few hours. Once the phenomenon is over, logarithmic returns usually remain in a smaller range. This is another reason for the value of the kurtosis being so large. Moreover, a seasonal peak in the winter was found for the case of Nord Pool, in contrast to the Californian electricity market, where the highest consumption is in summer. Since in the Nordic countries there is more need for heating in the winter season, the electricity prices are higher according to this. Also, with the frequencies of the largest price fluctuations for

every month, the peak in the winter was observable and confirmed, with a surprisingly high frequency also presented in October. In summer, the use of air conditioning is not as great as in California, which is why returns are less volatile in this period for the data considered. A detailed analysis of the autocorrelation function of the logarithmic returns was conducted. It was also found that there is not only a yearly rhythm but also a weekly as well as a daily one. The correlations between returns with 24 (day) and 168 (week) lags of difference are highly positively correlated. Furthermore, such a correlation lasts for almost all of the data points considered. This means that for $k = 24\tilde{k}$, $\tilde{k} \in \{1, 2, \dots, \tilde{n}\}$ the correlation was almost always positive and significant. The same consideration also holds for weekly correlation, indeed every 168 lags. In the autocorrelation within 24 lags, two relevant negative correlations were found at lags 6 and 18 respectively. The reason was that in an electricity market, returns are constantly switching from the transition to the stable phases and vice versa. It is highly probable that after 6 and 18 hours returns will have changed phase. Since returns in those phases have completely different behaviour, this explains the anti-correlation.

Another important feature of electricity and indeed the mean reversion was found thanks to the autocorrelation function of the lagged differences between returns. This autocorrelation decays very quickly and tends to zero, meaning that over the long-term the data considered will tend to its mean which is zero.

The decay of the autocorrelation function was also analysed. An analytic rate of approximately $-3/10$ was found. This is a rather slow decay and confirms the fact that the daily and weekly periodicities last over almost all the time considered. When only yearly autocorrelation functions were considered, decays in the range from $-3/5$ to $-2/5$ were found, which were indeed faster than the decay for the autocorrelation with n^* lags. This is because in the year only daily and weekly periodicities are accounted for. However, when considering correlations over several years, also a yearly periodicity is present, making the decay slower.

Finally, the laminar phases of the logarithmic returns and the corresponding probability density functions for the thresholds $\epsilon \in \{0.05, 0.1, 0.2, 0.3\}$ were computed. Such densities were expected to decay as a $-3/2$ power law over the increased length of time intervals. For thresholds 0.05 and 0.3, a too fast and too slow decay were found respectively. For a threshold between 0.1 and 0.2, the desired power law applies and the logarithmic returns of the spot prices from Nord Pool behave as the canonical theory of “on-off intermittency” states.

Appendix A

Programming Code in R

In this section, the programming code in R for figures, computations and tests is enlisted.

In L^AT_EX, the package called “listings” enables R codes to be displayed in a fashionable manner.¹

First of all, the working directory has to be set in the right place and the package to transfer the data from Excel to R has to be installed.

```
1 setwd("~/Dissertation Official")
2 require(XLConnect)
```

Logarithmic returns, spot prices and dates are extracted.

```
1 LogReturns.wb=loadWorkbook("Data.xlsx")
2 LogReturns.df=readWorksheet(LogReturns.wb,sheet="Data")
3 LogReturns=LogReturns.df$Lognormal>Returns
4 AbsLogReturns=abs(LogReturns)
5 n=length(LogReturns)
6 DatesLogReturns=LogReturns.df$Date
7 DatesSpotPrices=c(1999+1/(365.25*24), DatesLogReturns)
8 SpotPrices=rep(0,70752)
9 SpotPrices[1]=15.29
10 for(i in 2:70752)
11   SpotPrices[i]=SpotPrices[i-1]*exp(LogReturns[i-1])
12 rm(i)
```

Figure 2.1, which shows a time plot of the spot prices and logarithmic returns respectively, is displayed.

```
1 plot(DatesSpotPrices,SpotPrices,type="l", xlab="Time", ylab="S(t)")
2 plot(DatesLogReturns,LogReturns,type="l", xlab="Time", ylab="r(t)")
```

Spot prices and logarithmic returns are converted to time series and Figure 2.2 is displayed.

```
1 SpotPrices.ts=ts(SpotPrices,start=1999+1/(365.25*24), deltat=1/(365.25*24))
2 LogReturns.ts1=ts(LogReturns,start=1999+2/(365.25*24), deltat=1/(365.25*24))
3 LogReturns.ts2=ts(LogReturns,start=1,deltat=1)
4 plot(decompose(SpotPrices.ts))
```

¹The settings for such a display are taken from the following [Stack Overflow](#) link.

Figure C.1 which shows the yearly trend from the Appendix is displayed.

```
1 WeeklyTrend=filter(SpotPrices.ts,filter=rep(1/168,168))
2 MonthlyTrend=filter(SpotPrices.ts,filter=rep(1/(168*4),168*4))
3 YearlyTrend=filter(SpotPrices.ts,filter=rep(1/8766,8766))
4 plot(WeeklyTrend,ylab="m(t)",col="red")
5 lines(MonthlyTrend,col="green")
6 lines(YearlyTrend,col="purple")
7 legend(1999.1,110,legend=c("Daily trend","Weekly trend","Yearly trend"),col=c("red","green","purple"),lty=c(1,1,1),
      bty="n")
```

Spot prices for one week are displayed in Figure 2.3.

```
1 OneWeek.ts=window(SpotPrices.ts,2002.420,2002.439)
2 OneWeek=unclass(OneWeek.ts)
3 Pos1=(length(OneWeek)/7)*(1:7)-(length(OneWeek)/7)/2
4 plot(OneWeek,type="l",xaxt="n",col="blue",xlab="Days",ylab="S(t)")
5 axis(1,Pos1,labels=c("Mon","Tue","Wed","Thu","Fri","Sat","Sun"))
```

Yearly logarithmic returns from original time series are extracted.

```
1 LogReturns1999=window(LogReturns.ts1, start=1999.00023, end=1999.99989)
2 LogReturns2000=window(LogReturns.ts1, start=2000, end=2000.99989)
3 LogReturns2001=window(LogReturns.ts1, start=2001, end=2001.99989)
4 LogReturns2002=window(LogReturns.ts1, start=2002, end=2002.99989)
5 LogReturns2003=window(LogReturns.ts1, start=2003, end=2003.99989)
6 LogReturns2004=window(LogReturns.ts1, start=2004, end=2004.99989)
7 LogReturns2005=window(LogReturns.ts1, start=2005, end=2005.99989)
8 LogReturns2006=window(LogReturns.ts1, start=2006, end=2006.99989)
9 LogReturns2007=window(LogReturns.ts1, start=2007, end=2007.07118)
```

The empirical descriptive statistics for the logarithmic returns are computed.

```
1 Mu=mean(LogReturns)
2 Variance=var(LogReturns)
3 StandardDeviation=sd(LogReturns)
4 require(moments)
5 Skewness=skewness(LogReturns)
6 Kurtosis=kurtosis(LogReturns)
```

From the above descriptive statistics, it is possible to calculate the test statistics for the mean, skewness and kurtosis described and to apply the Jarque-Bera test.

```
1 TStatMean=(sqrt(n)*Mu)/StandardDeviation
2 TStatSkewness=Skewness/(sqrt(6/n))
3 TStatKurtosis=(Kurtosis-3)/(sqrt(24/n))
4 jarque.test(LogReturns)
```

Figure 3.1, which shows the distribution of the logarithmic returns against a normal distribution, is displayed.

```
1 plot(density(LogReturns.ts1),lwd=2.5,xlab="Logarithmic Returns",main="",xlim=c(-0.2,0.2),col="blue4")
2 x=seq(-0.2,0.2,length=100)
3 y=dnorm(x,Mu,StandardDeviation)
4 lines(x,y,lwd=2.5,col="darkgreen")
5 legend(-0.2,25, legend=c("Logarithmic returns","Normal distribution"),lty=c(1,1), lwd=c(2.5,2.5),col=c("blue4","darkgreen"))
```

Daily logarithmic returns from original time series are extracted.

```

1 DailyLogReturns1999=LogReturns1999[6+24*(0:364)]
2 DailyLogReturns2000=LogReturns2000[8+24*(0:364)]
3 DailyLogReturns2001=LogReturns2001[8+24*(0:364)]
4 DailyLogReturns2002=LogReturns2002[8+24*(0:364)]
5 DailyLogReturns2003=LogReturns2003[8+24*(0:364)]
6 DailyLogReturns2004=LogReturns2004[8+24*(0:364)]
7 DailyLogReturns2005=LogReturns2005[8+24*(0:364)]
8 DailyLogReturns2006=LogReturns2006[8+24*(0:364)]

```

Mean, standard deviation and test statistics for every year are computed.

```

1 MeanOfYears=c(mean(LogReturns1999),mean(LogReturns2000),mean(LogReturns2001),mean(LogReturns2002),mean(
  LogReturns2003),mean(LogReturns2004),mean(LogReturns2005),mean(LogReturns2006),mean(LogReturns2007))
2 StandardDeviationOfYears=c(sd(LogReturns1999),sd(LogReturns2000),sd(LogReturns2001),sd(LogReturns2002),sd(
  LogReturns2003),sd(LogReturns2004),sd(LogReturns2005),sd(LogReturns2006),sd(LogReturns2007))
3 TStatMeanOfYears=sqrt(n)*(MeanOfYears/StandardDeviationOfYears)

```

Expectation, variance, standard deviation, cardinality and test statistic for “differences sign test” with n^* observations are computed.

```

1 ExpectationOfCardinality=1/2*(n-1)
2 VarianceOfCardinality=1/12*(n+1)
3 StandardDeviationOfCardinality=sqrt(VarianceOfCardinality)
4 DifferencesOfLogReturns=LogReturns[2:70751]-LogReturns[1:70750]
5 #the elements of the set S are found
6 Cardinality=length(DifferencesOfLogReturns[DifferencesOfLogReturns>0])
7 StatCardinality=abs(Cardinality-ExpectationOfCardinality)/StandardDeviationOfCardinality

```

Expectation, variance and standard deviation for “differences sign test” with 365 observations are computed.

```

1 Days=365
2 ExpectationOfCardinality365=1/2*(Days-1)
3 VarianceOfCardinality365=1/12*(Days+1)
4 StandardDeviationOfCardinality365=sqrt(VarianceOfCardinality365)

```

The differences of the daily logarithmic returns for every year are computed.

```

1 DifferencesOfLogReturns1999=DailyLogReturns1999[2:365]-DailyLogReturns1999[1:364]
2 DifferencesOfLogReturns2000=DailyLogReturns2000[2:365]-DailyLogReturns2000[1:364]
3 DifferencesOfLogReturns2001=DailyLogReturns2001[2:365]-DailyLogReturns2001[1:364]
4 DifferencesOfLogReturns2002=DailyLogReturns2002[2:365]-DailyLogReturns2002[1:364]
5 DifferencesOfLogReturns2003=DailyLogReturns2003[2:365]-DailyLogReturns2003[1:364]
6 DifferencesOfLogReturns2004=DailyLogReturns2004[2:365]-DailyLogReturns2004[1:364]
7 DifferencesOfLogReturns2005=DailyLogReturns2005[2:365]-DailyLogReturns2005[1:364]
8 DifferencesOfLogReturns2006=DailyLogReturns2006[2:365]-DailyLogReturns2006[1:364]

```

Cardinality and test statistics for every year are calculated.

```

1 Cardinality1999=length(DifferencesOfLogReturns1999[DifferencesOfLogReturns1999>0])
2 Cardinality2000=length(DifferencesOfLogReturns2000[DifferencesOfLogReturns2000>0])
3 Cardinality2001=length(DifferencesOfLogReturns2001[DifferencesOfLogReturns2001>0])
4 Cardinality2002=length(DifferencesOfLogReturns2002[DifferencesOfLogReturns2002>0])
5 Cardinality2003=length(DifferencesOfLogReturns2003[DifferencesOfLogReturns2003>0])
6 Cardinality2004=length(DifferencesOfLogReturns2004[DifferencesOfLogReturns2004>0])
7 Cardinality2005=length(DifferencesOfLogReturns2005[DifferencesOfLogReturns2005>0])
8 Cardinality2006=length(DifferencesOfLogReturns2006[DifferencesOfLogReturns2006>0])
9 CardinalityOfYears=c(Cardinality1999, Cardinality2000, Cardinality2001, Cardinality2002, Cardinality2003,
  Cardinality2004, Cardinality2005, Cardinality2006)
10 StatCardinalityOfYears=abs(CardinalityOfYears-ExpectationOfCardinality365)/StandardDeviationOfCardinality365

```

Autocorrelation function with n^* lags and the corresponding absolute value is computed. In the same moment, Figure 4.1 is displayed.

```
1 ACFList=acf(LogReturns.ts1,lag=70751)
2 ACF=ACFList$acf
3 AbsACF=abs(ACF)
4 DailyAbsACF=AbsACF[1+24*(0:2947)]
```

Figure 4.2, which shows the autocorrelation of the logarithmic returns for different lags, is displayed.

```
1 par(mfrow=c(2,2))
2 #one day
3 acf(LogReturns.ts2,24,main="(i)")
4 #one week
5 acf(LogReturns.ts2,168,main="(ii)")
6 #two weeks
7 acf(LogReturns.ts2,168*2,main="(iii)")
8 #one month
9 acf(LogReturns.ts2,168*4,main="(iv)")
10 par(mfrow=c(1,1))
```

Figure 4.3, which shows mean reversion, is displayed.

```
1 DailyLogReturns=LogReturns.ts2[24*(0:2947)]
2 #the following command is used to find z(t)
3 DiffLogReturns=diff(DailyLogReturns,lag=7)
4 DiffAcf.acf=acf(DiffLogReturns,lag=100,plot=FALSE)
5 DiffAcf=DiffAcf.acf$acf
6 Pos2=7*(0:14)
7 plot(DiffAcf,type="o",lty=1,xlab="Lag (days)",ylab="Lagged difference of ACF",xaxt="n")
8 axis(1,1+Pos2,labels=Pos2)
9 abline(0,0)
10 #95% confidence interval for the autocorrelation function
11 ConfInt=1.96/sqrt(length(DiffLogReturns))
12 abline(ConfInt,0,lty=2,col="blue")
13 abline(-ConfInt,0,lty=2,col="blue")
14 abline(v=1+Pos2,lty=8,col="grey")
15 abline(h=c(1,0.5,-0.5),lty=8,col="grey")
```

The autocorrelation function for every year is calculated.

```
1 ACF1999.acf=acf(LogReturns1999,lag=8763,plot=FALSE)
2 ACF1999=ACF1999.acf$acf
3 ACF2000.acf=acf(LogReturns2000,lag=8766,plot=FALSE)
4 ACF2000=ACF2000.acf$acf
5 ACF2001.acf=acf(LogReturns2001,lag=8766,plot=FALSE)
6 ACF2001=ACF2001.acf$acf
7 ACF2002.acf=acf(LogReturns2002,lag=8766,plot=FALSE)
8 ACF2002=ACF2002.acf$acf
9 ACF2003.acf=acf(LogReturns2003,lag=8766,plot=FALSE)
10 ACF2003=ACF2003.acf$acf
11 ACF2004.acf=acf(LogReturns2004,lag=8766,plot=FALSE)
12 ACF2004=ACF2004.acf$acf
13 ACF2005.acf=acf(LogReturns2005,lag=8766,plot=FALSE)
14 ACF2005=ACF2005.acf$acf
15 ACF2006.acf=acf(LogReturns2006,lag=8766,plot=FALSE)
16 ACF2006=ACF2006.acf$acf
```

Decays of several autocorrelation functions are displayed in the Appendix.

```

1 AllACF=cbind(ACF1999[1+24*(0:365)],ACF2000[1+24*(0:365)],ACF2001[1+24*(0:365)],ACF2002[1+24*(0:365)],ACF2003[1+24*(0:365)],ACF2004[1+24*(0:365)],ACF2005[1+24*(0:365)],ACF2006[1+24*(0:365)])
2 x.name = c("ACF99","ACF00","ACF01","ACF02","ACF03","ACF04","ACF05","ACF06")
3 colnames(AllACF) = c(x.name[1:4], x.name[5:8])
4 plot.ts(AllACF, main="")
5 plot.ts(AllACF, main="",log="xy")

```

The independent and dependent variables for linear regression are computed.

```

1 #for the autocorrelation with 2947 days (complete set), the cutoff is 1700 and this is why 1:1700 is considered
2 RegXComplete=log(1:1700)
3 RegX=log(1:200)
4 RegACF=log(DailyAbsACF[1:1700])
5 #cutoff for yearly autocorrelation is approximately 199
6 RegACF1999=log(ACF1999[1+24*(0:199)])
7 RegACF2000=log(ACF2000[1+24*(0:199)])
8 RegACF2001=log(ACF2001[1+24*(0:199)])
9 RegACF2002=log(ACF2002[1+24*(0:199)])
10 RegACF2003=log(ACF2003[1+24*(0:199)])
11 RegACF2004=log(ACF2004[1+24*(0:199)])
12 RegACF2005=log(ACF2005[1+24*(0:199)])
13 RegACF2006=log(ACF2006[1+24*(0:199)])

```

The linear regression for the decay of the autocorrelation functions is performed.

```

1 #linear regression for autocorrelation
2 FitACF=lm(RegACF~RegXComplete)
3 FitACF1999=lm(RegACF1999~RegX)
4 FitACF2000=lm(RegACF2000~RegX)
5 FitACF2001=lm(RegACF2001~RegX)
6 FitACF2002=lm(RegACF2002~RegX)
7 FitACF2003=lm(RegACF2003~RegX)
8 FitACF2004=lm(RegACF2004~RegX)
9 FitACF2005=lm(RegACF2005~RegX)
10 FitACF2006=lm(RegACF2006~RegX)
11 #coefficients are stored in a vector, where intercept and regression are first and second elements respectively
12 CoeffACF=FitACF$coefficients
13 CoeffACF1999=FitACF1999$coefficients
14 CoeffACF2000=FitACF2000$coefficients
15 CoeffACF2001=FitACF2001$coefficients
16 CoeffACF2002=FitACF2002$coefficients
17 CoeffACF2003=FitACF2003$coefficients
18 CoeffACF2004=FitACF2004$coefficients
19 CoeffACF2005=FitACF2005$coefficients
20 CoeffACF2006=FitACF2006$coefficients

```

The decay for the autocorrelation function with n^* lags and the corresponding power law is displayed in Figure 4.4.

```

1 xComplete1=1:length(DailyAbsACF)
2 yComplete1=exp(CoeffACF[1])*xComplete1^(CoeffACF[2])
3 plot(DailyAbsACF,type="l",col="blue4",ylab="Absolute value of ACF",xlab="Days")
4 lines(yComplete1,col="violet")
5 legend(2000,1,c("ACF","Power law of -3/10"),lwd=2.5,col=c("blue4","violet"))
6 xComplete2=1:1700
7 yComplete2=exp(CoeffACF[1])*xComplete2^(CoeffACF[2])
8 plot(DailyAbsACF[1:1700],type="l",log="xy",ylim=c(0.06,1),col="blue4",ylab="Absolute value of ACF",xlab="Days")
9 lines(yComplete2,col="violet")
10 legend(150,1,c("ACF","Power law of -3/10"),lwd=2.5,col=c("blue4","violet"))

```

The decay of the autocorrelation functions for every year is analysed and displayed in the Appendix, with the corresponding power law.

```

1 #to plot the linear regression, intercept and slope are used and can be found in the tables
2 par(mfrow=c(1,2))
3 x=1:200
4 plot(ACF1999[1+24*(0:199)],type="l",xlab="Lag (Days)", ylab="ACF 1999",log="xy",ylim=c(0.1,1))
5 y=exp(CoeffACF1999[1])*x^(CoeffACF1999[2])
6 lines(y,col="red")
7 plot(ACF2000[1+24*(0:199)],type="l",xlab="Lag (Days)", ylab="ACF 2000",log="xy",ylim=c(0.1,1.2))
8 y=exp(CoeffACF2000[1])*x^(CoeffACF2000[2])
9 lines(y,col="red")
10 plot(ACF2001[1+24*(0:199)],type="l",xlab="Lag (Days)", ylab="ACF 2001",log="xy",ylim=c(0.05,1.2))
11 y=exp(CoeffACF2001[1])*x^(CoeffACF2001[2])
12 lines(y,col="red")
13 plot(ACF2002[1+24*(0:199)],type="l",xlab="Lag (Days)", ylab="ACF 2002",log="xy",ylim=c(0.05,1.5))
14 y=exp(CoeffACF2002[1])*x^(CoeffACF2002[2])
15 lines(y,col="red")
16 plot(ACF2003[1+24*(0:199)],type="l",xlab="Lag (Days)", ylab="ACF 2003",log="xy",ylim=c(0.05,2))
17 y=exp(CoeffACF2003[1])*x^(CoeffACF2003[2])
18 lines(y,col="red")
19 plot(ACF2004[1+24*(0:199)],type="l",xlab="Lag (Days)", ylab="ACF 2004",log="xy",ylim=c(0.05,1))
20 y=exp(CoeffACF2004[1])*x^(CoeffACF2004[2])
21 lines(y,col="red")
22 plot(ACF2005[1+24*(0:199)],type="l",xlab="Lag (Days)", ylab="ACF 2005",log="xy",ylim=c(0.05,1))
23 y=exp(CoeffACF2005[1])*x^(CoeffACF2005[2])
24 lines(y,col="red")
25 plot(ACF2006[1+24*(0:199)],type="l",xlab="Lag (Days)", ylab="ACF 2006",log="xy",ylim=c(0.05,1))
26 y=exp(CoeffACF2006[1])*x^(CoeffACF2006[2])
27 lines(y,col="red")
28 par(mfrow=c(1,1))

```

To divide the year into the respective months, the following code is used.

```

1 split1999=split(LogReturns1999,ceiling(seq_along(LogReturns1999)/(length(LogReturns1999)/12)))
2 split2000=split(LogReturns2000,ceiling(seq_along(LogReturns2000)/(length(LogReturns2000)/12)))
3 split2001=split(LogReturns2001,ceiling(seq_along(LogReturns2001)/(length(LogReturns2001)/12)))
4 split2002=split(LogReturns2002,ceiling(seq_along(LogReturns2002)/(length(LogReturns2002)/12)))
5 split2003=split(LogReturns2003,ceiling(seq_along(LogReturns2003)/(length(LogReturns2003)/12)))
6 split2004=split(LogReturns2004,ceiling(seq_along(LogReturns2004)/(length(LogReturns2004)/12)))
7 split2005=split(LogReturns2005,ceiling(seq_along(LogReturns2005)/(length(LogReturns2005)/12)))
8 split2006=split(LogReturns2006,ceiling(seq_along(LogReturns2006)/(length(LogReturns2006)/12)))

```

Logarithmic returns are grouped for every month of the year.

```

1 January=c(split1999$'1',split2000$'1',split2001$'1',split2002$'1',split2003$'1',split2004$'1',split2005$'1',
2           split2006$'1')
3 February=c(split1999$'2',split2000$'2',split2001$'2',split2002$'2',split2003$'2',split2004$'2',split2005$'2',
4           split2006$'2')
5 March=c(split1999$'3',split2000$'3',split2001$'3',split2002$'3',split2003$'3',split2004$'3',split2005$'3',split2006$
6           $'3')
7 April=c(split1999$'4',split2000$'4',split2001$'4',split2002$'4',split2003$'4',split2004$'4',split2005$'4',split2006$
8           $'4')
9 May=c(split1999$'5',split2000$'5',split2001$'5',split2002$'5',split2003$'5',split2004$'5',split2005$'5',split2006$
10          $'5')
11 June=c(split1999$'6',split2000$'6',split2001$'6',split2002$'6',split2003$'6',split2004$'6',split2005$'6',split2006$
12          $'6')
13 July=c(split1999$'7',split2000$'7',split2001$'7',split2002$'7',split2003$'7',split2004$'7',split2005$'7',split2006$
14          $'7')
15 August=c(split1999$'8',split2000$'8',split2001$'8',split2002$'8',split2003$'8',split2004$'8',split2005$'8',
16           split2006$'8')
17 September=c(split1999$'9',split2000$'9',split2001$'9',split2002$'9',split2003$'9',split2004$'9',split2005$'9',
18           split2006$'9')

```

```

1 October=c(split1999$'10',split2000$'10',split2001$'10',split2002$'10',split2003$'10',split2004$'10',split2005$'10',
  split2006$'10')
2 November=c(split1999$'11',split2000$'11',split2001$'11',split2002$'11',split2003$'11',split2004$'11',split2005$
  '11',split2006$'11')
3 December=c(split1999$'12',split2000$'12',split2001$'12',split2002$'12',split2003$'12',split2004$'12',split2005$
  '12',split2006$'12')

```

Absolute values are computed.

```

1 AbsJanuary=abs(January)
2 AbsFebruary=abs(February)
3 AbsMarch=abs(March)
4 AbsApril=abs(April)
5 AbsMay=abs(May)
6 AbsJune=abs(June)
7 AbsJuly=abs(July)
8 AbsAugust=abs(August)
9 AbsSeptember=abs(September)
10 AbsOctober=abs(October)
11 AbsNovember=abs(November)
12 AbsDecember=abs(December)

```

The 100 biggest price fluctuations are calculated.

```

1 SortAbsLogRet=sort(AbsLogReturns,decreasing=TRUE)
2 BiggestAbsReturns=SortAbsLogRet[1:100]
3 MinOfBiggestAbsReturns=BiggestAbsReturns[100]

```

To check how many large returns are in each month, the following code is used.

```

1 AbsLargestJanuary=AbsJanuary[AbsJanuary>MinOfBiggestAbsReturns]
2 AbsLargestFebruary=AbsFebruary[AbsFebruary>MinOfBiggestAbsReturns]
3 AbsLargestMarch=AbsMarch[AbsMarch>MinOfBiggestAbsReturns]
4 AbsLargestApril=AbsApril[AbsApril>MinOfBiggestAbsReturns]
5 AbsLargestMay=AbsMay[AbsMay>MinOfBiggestAbsReturns]
6 AbsLargestJune=AbsJune[AbsJune>MinOfBiggestAbsReturns]
7 AbsLargestJuly=AbsJuly[AbsJuly>MinOfBiggestAbsReturns]
8 AbsLargestAugust=AbsAugust[AbsAugust>MinOfBiggestAbsReturns]
9 AbsLargestSeptember=AbsSeptember[AbsSeptember>MinOfBiggestAbsReturns]
10 AbsLargestOctober=AbsOctober[AbsOctober>MinOfBiggestAbsReturns]
11 AbsLargestNovember=AbsNovember[AbsNovember>MinOfBiggestAbsReturns]
12 AbsLargestDecember=AbsDecember[AbsDecember>MinOfBiggestAbsReturns]

```

The number of extreme observations per month are computed.

```

1 LAbsJanuary=length(AbsLargestJanuary)
2 LAbsFebruary=length(AbsLargestFebruary)
3 LAbsMarch=length(AbsLargestMarch)
4 LAbsApril=length(AbsLargestApril)
5 LAbsMay=length(AbsLargestMay)
6 LAbsJune=length(AbsLargestJune)
7 LAbsJuly=length(AbsLargestJuly)
8 LAbsAugust=length(AbsLargestAugust)
9 LAbsSeptember=length(AbsLargestSeptember)
10 LAbsOctober=length(AbsLargestOctober)
11 LAbsNovember=length(AbsLargestNovember)
12 LAbsDecember=length(AbsLargestDecember)
13 LAbsMonths=c(LAbsJanuary,LAbsFebruary,LAbsMarch,LAbsApril,LAbsMay,LAbsJune,LAbsJuly,LAbsAugust,LAbsSeptember,
  LAbsOctober,LAbsNovember,LAbsDecember)

```

Finally, the frequencies of the largest 100 observations are displayed with respect to the months. This corresponds to Figure 5.1.

```
1 LabMonths=c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")
2 library(lattice)
3 dotplot(LabMonths~factor(LabMonths, levels = LabMonths), axes=FALSE, xlab="Months", ylab="Frequency")
```

The laminar phases for $\epsilon \in \{0.05, 0.1, 0.2, 0.3\}$ are calculated.

```
1 AbsLogReturnsThreshold0.05=AbsLogReturns
2 AbsLogReturnsThreshold0.05[AbsLogReturnsThreshold0.05<0.05]=0
3 AbsLogReturnsThreshold0.1=AbsLogReturns
4 AbsLogReturnsThreshold0.1[AbsLogReturnsThreshold0.1<0.1]=0
5 AbsLogReturnsThreshold0.2=AbsLogReturns
6 AbsLogReturnsThreshold0.2[AbsLogReturnsThreshold0.2<0.2]=0
7 AbsLogReturnsThreshold0.3=AbsLogReturns
8 AbsLogReturnsThreshold0.3[AbsLogReturnsThreshold0.3<0.3]=0
9 #ROEV stands for "runs of equal values"
10 ROEV0.05=rle(AbsLogReturnsThreshold0.05)
11 LaminarPhases0.05=ROEV0.05$lengths[ROEV0.05$values==0]
12 ROEV0.1=rle(AbsLogReturnsThreshold0.1)
13 LaminarPhases0.1=ROEV0.1$lengths[ROEV0.1$values==0]
14 ROEV0.2=rle(AbsLogReturnsThreshold0.2)
15 LaminarPhases0.2=ROEV0.2$lengths[ROEV0.2$values==0]
16 ROEV0.3=rle(AbsLogReturnsThreshold0.3)
17 LaminarPhases0.3=ROEV0.3$lengths[ROEV0.3$values==0]
```

The plots of the intermittency behaviour for several ϵ in the Appendix are displayed.

```
1 plot(DatesLogReturns, AbsLogReturnsThreshold0.05, type="l", xlab="Time", ylab="r*(t)")
2 abline(0.05, 0, col="orangered", lty=2, lwd=1)
3 plot(DatesLogReturns, AbsLogReturnsThreshold0.1, type="l", xlab="Time", ylab="r*(t)")
4 abline(0.1, 0, col="orangered", lty=2, lwd=1)
5 plot(DatesLogReturns, AbsLogReturnsThreshold0.2, type="l", xlab="Time", ylab="r*(t)")
6 abline(0.2, 0, col="orangered", lty=2, lwd=1)
7 plot(DatesLogReturns, AbsLogReturnsThreshold0.3, type="l", xlab="Time", ylab="r*(t)")
8 abline(0.3, 0, col="orangered", lty=2, lwd=1)
```

The Kernel densities are calculated.

```
1 dens0.05=density(LaminarPhases0.05)
2 xdens0.05=dens0.05$x[6:512]
3 ydens0.05=dens0.05$y[6:512]
4 dens0.1=density(LaminarPhases0.1)
5 xdens0.1=dens0.1$x[4:512]
6 ydens0.1=dens0.1$y[4:512]
7 dens0.2=density(LaminarPhases0.2)
8 xdens0.2=dens0.2$x[10:512]
9 ydens0.2=dens0.2$y[10:512]
10 dens0.3=density(LaminarPhases0.3)
11 xdens0.3=dens0.3$x[14:512]
12 ydens0.3=dens0.3$y[14:512]
```

The probability density functions of the laminar phases for different thresholds are coerced to data frames and adjusted from the previously computed Kernel densities.

```
1 densities0.05=data.frame(x=xdens0.05, y=ydens0.05)
2 densities0.1=data.frame(x=xdens0.1, y=ydens0.1)
3 densities0.2=data.frame(x=xdens0.2, y=ydens0.2)
4 densities0.3=data.frame(x=xdens0.3, y=ydens0.3)
5 NonZeros0.05=apply(densities0.05, 1, function(row) all(row !=0 ))
6 NonZeros0.1=apply(densities0.1, 1, function(row) all(row !=0 ))
```



```

1 NonZeros0.2=apply(densities0.2, 1, function(row) all(row !=0 ))
2 NonZeros0.3=apply(densities0.3, 1, function(row) all(row !=0 ))
3 #adjusted densities are without zeros.
4 AdjDens0.05=densities0.05[NonZeros0.05,]
5 AdjDens0.1=densities0.1[NonZeros0.1,]
6 AdjDens0.2=densities0.2[NonZeros0.2,]
7 AdjDens0.3=densities0.3[NonZeros0.3,]

```

A linear regression of the logarithm of probability density against the logarithm of lengths of laminar phases is performed.

```

1 xpdf0.05=log(AdjDens0.05$x[1:81])
2 ypdf0.05=log(AdjDens0.05$y[1:81])
3 xpdf0.1=log(AdjDens0.1$x[1:88])
4 ypdf0.1=log(AdjDens0.1$y[1:88])
5 xpdf0.2=log(AdjDens0.2$x[1:137])
6 ypdf0.2=log(AdjDens0.2$y[1:137])
7 xpdf0.3=log(AdjDens0.3$x[1:137])
8 ypdf0.3=log(AdjDens0.3$y[1:137])
9 fit0.05=lm(ypdf0.05~xpdf0.05)
10 fit0.1=lm(ypdf0.1~xpdf0.1)
11 fit0.2=lm(ypdf0.2~xpdf0.2)
12 fit0.3=lm(ypdf0.3~xpdf0.3)
13 Coeff0.05=fit0.05$coefficients
14 Coeff0.1=fit0.1$coefficients
15 Coeff0.2=fit0.2$coefficients
16 Coeff0.3=fit0.3$coefficients

```

Figure 5.2 which shows the decays for the laminar phases of different thresholds is plotted on a double logarithmic scale.

```

1 #the final element of the plot (e.g. 88) is chosen in a way that it is easier to observe the several decays for
  different values of thresholds
2 plot(AdjDens0.1$x[1:88], AdjDens0.1$y[1:88], log="xy", col="green", type="o", xlim=c(AdjDens0.3$x[1], 1300), xlab="Length
  of laminar phases (hours)", ylab="Probability density")
3 lines(AdjDens0.05$x[1:81], AdjDens0.1$y[1:81], col="blue", type="o")
4 lines(AdjDens0.2$x[1:137], AdjDens0.2$y[1:137], col="red", type="o")
5 y=exp(Coeff0.2[1])*(AdjDens0.2$x[1:137])^(-1.5)
6 lines(AdjDens0.2$x[1:137], y, col="red")
7 lines(AdjDens0.3$x[1:137], AdjDens0.3$y[1:137], col="purple", type="o")
8 legend(300, 0.05, legend=c(expression(paste(epsilon, " = ", 0.05)), expression(paste(epsilon, " = ", 0.1)), expression(
  paste(epsilon, " = ", 0.2)), expression(paste(epsilon, " = ", 0.3))), lty=c(1, 1, 1, 1), col=c("blue", "green", "red", "
  purple"), bty="n", pch=c(1, 1, 1, 1))
9 #this way, all decays can be displayed in one figure, otherwise, there would be too many fluctuations after these
  values and it would be very difficult to observe a linear plot on a double logarithmic scale

```

Appendix B

Additional Tables

Years	Mean	Standard deviation	Test statistic
1999	-6.274677e-06	0.04410185	-0.03784433
2000	1.861529e-05	0.08013722	0.06178768
2001	5.008862e-05	0.08593348	0.15503970
2002	1.304966e-04	0.04670245	0.74323442
2003	-1.208470e-04	0.04693618	-0.68484830
2004	-2.290084e-05	0.04718298	-0.12910179
2005	4.758475e-05	0.05374739	0.23549226
2006	-2.727668e-05	0.05067206	-0.14318229
2007	-1.300494e-04	0.07196283	-0.48069158

Table B.1: Mean, standard deviation and test statistic for every year of the period considered.

Years	Cardinality	Test statistic
1999	182	0
2000	183	0.1810715
2001	193	1.9917864
2002	178	0.7242860
2003	183	0.1810715
2004	183	0.1810715
2005	182	0
2006	174	1.4485719

Table B.2: Results of the “differences sign test” for every year.

Years	α_0	α_1
1999	-0.08039349	-0.39247405
2000	0.1404176	-0.3852416
2001	0.1042413	-0.5034610
2002	0.4541387	-0.4879907
2003	0.5960216	-0.5884822
2004	-0.4761558	-0.4340570
2005	-0.2163494	-0.4558454
2006	-0.2335309	-0.4187047

Table B.3: Intercepts and analytic rates of the decay for the autocorrelation function of every year, found by a linear regression.

Appendix C

Additional Explanatory Figures

C.1 Daily, weekly and yearly trend

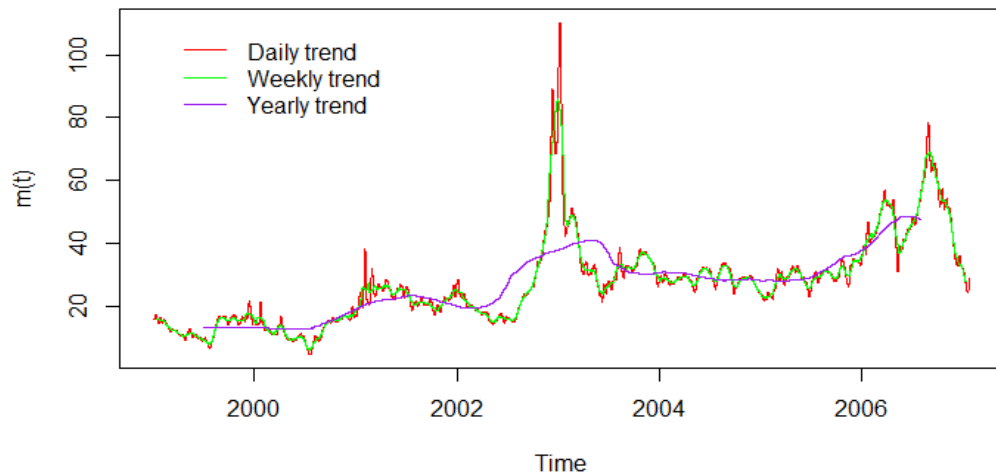


Figure C.1: Daily, weekly and yearly trends of spot prices computed by a moving average.

C.2 Decays of autocorrelation functions

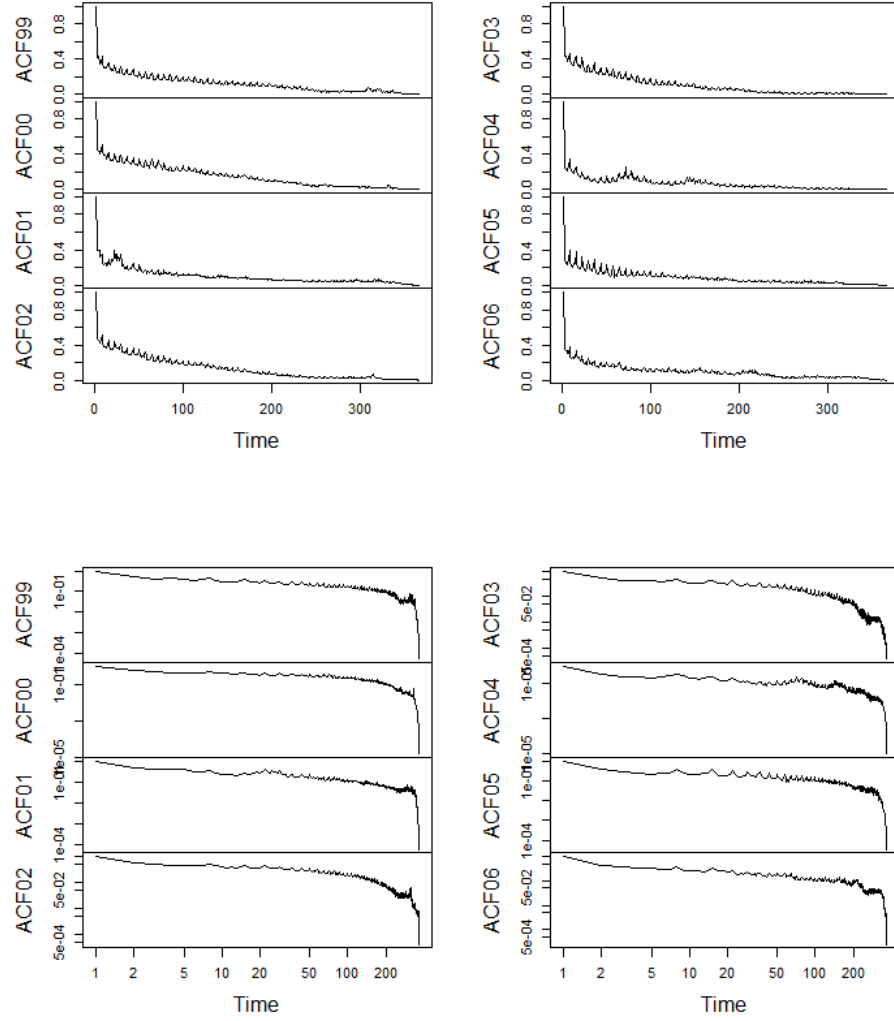


Figure C.2: Daily autocorrelation function for every year in both normal (top) and double logarithmic (bottom) plots. Such a display can be found also in [\[10\]](#) page 408.

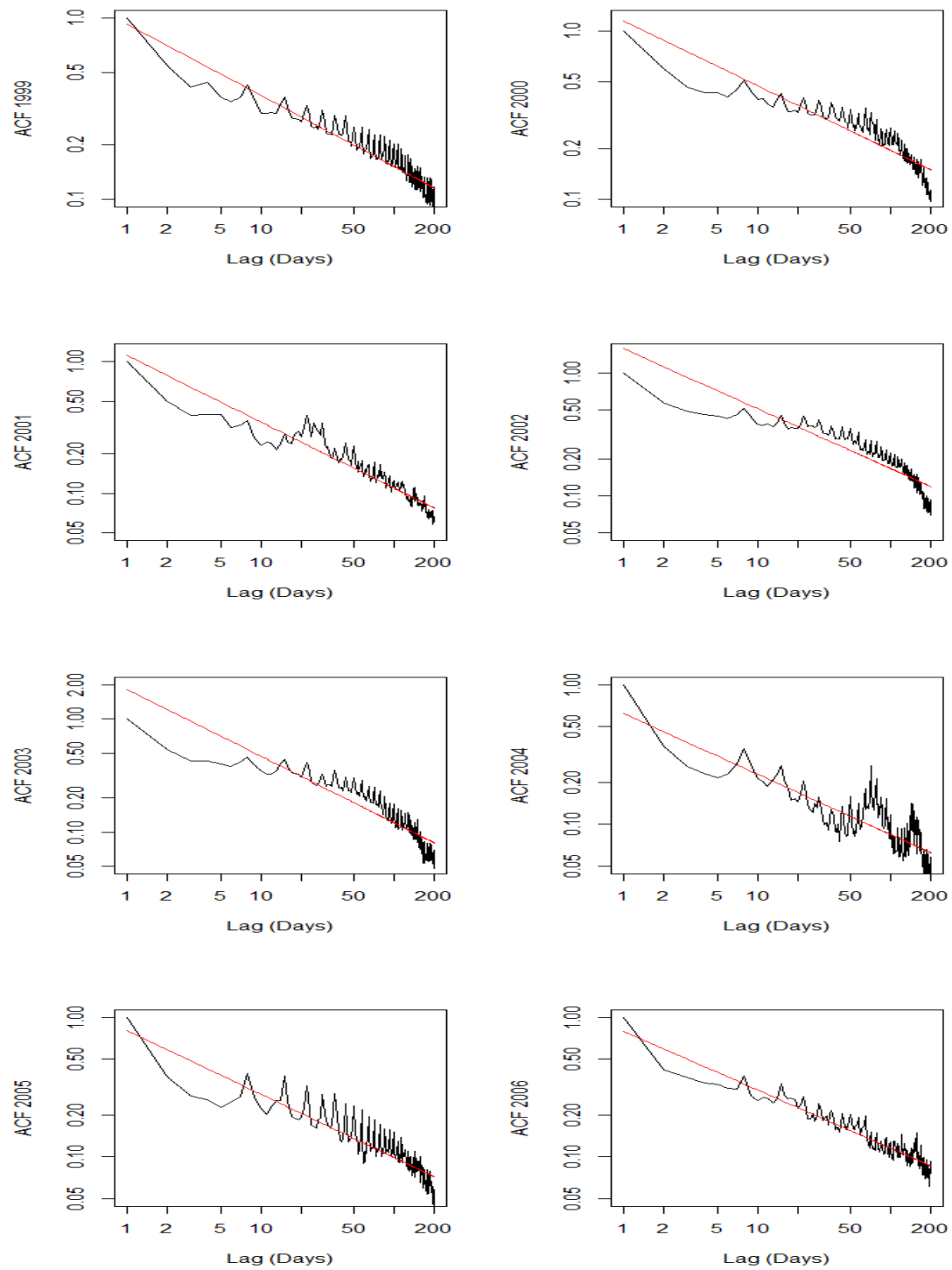


Figure C.3: Decay for the autocorrelation for every year from 1999 to 2006 with respective power laws.

C.3 On-off intermittency for different thresholds

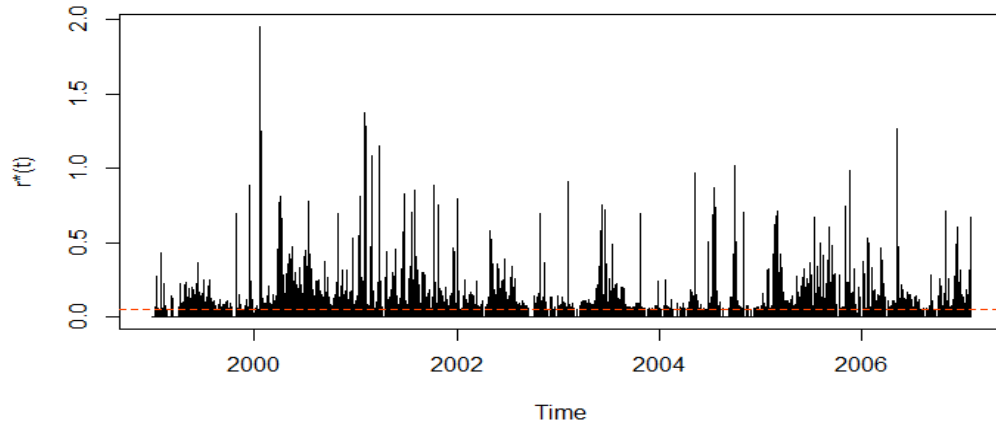


Figure C.4: Intermittent behaviour of logarithmic returns with n^* observation with $\epsilon = 0.05$.

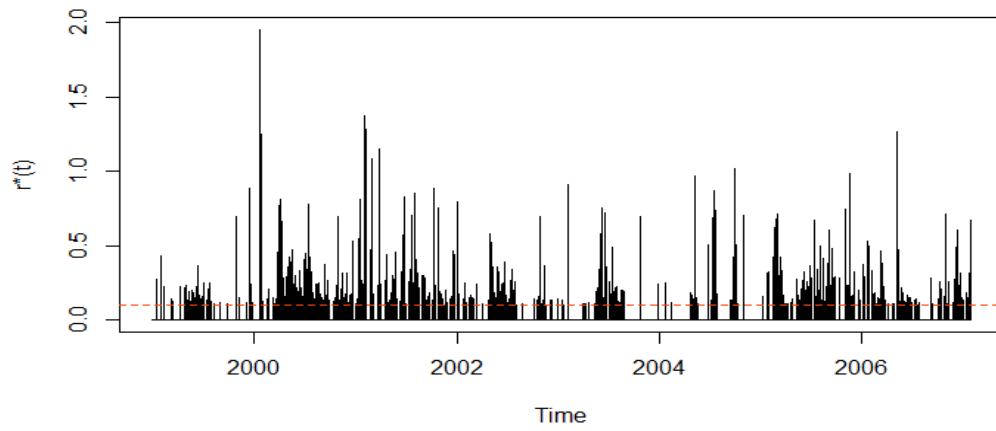


Figure C.5: Intermittent behaviour of logarithmic returns with n^* observation with $\epsilon = 0.1$.

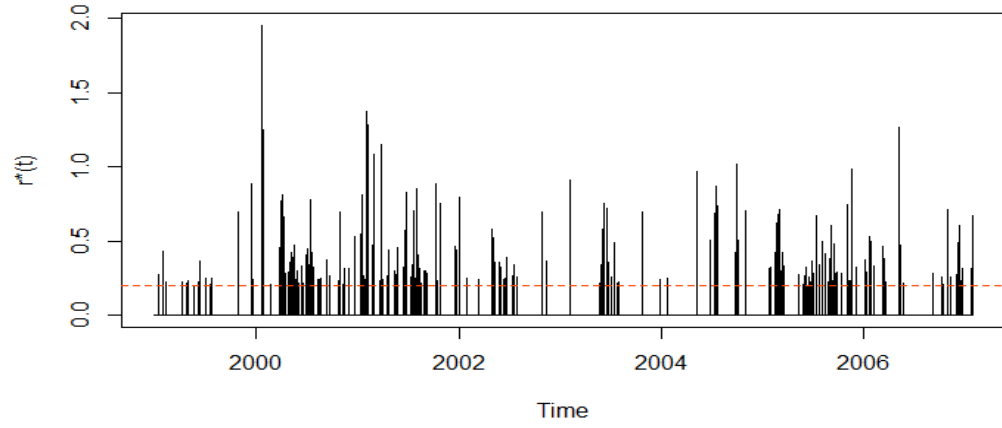


Figure C.6: Intermittent behaviour of logarithmic returns with n^* observation with $\epsilon = 0.2$.

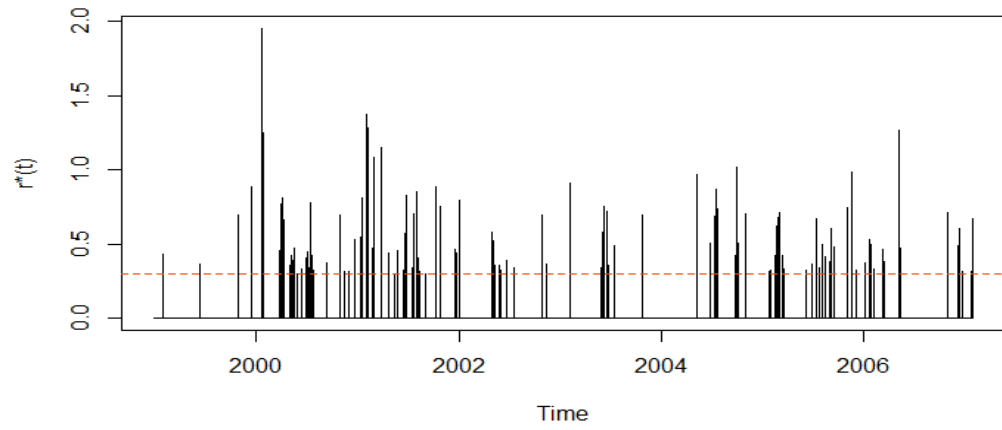


Figure C.7: Intermittent behaviour of logarithmic returns with n^* observation with $\epsilon = 0.3$.

Bibliography

- [1] Byström H. N. E., *Extreme value theory and extremely large electricity price changes*, Int. Rev. Econ. and Finance 14 (2005).
- [2] Cowpertwait P. S. P., *Introductory time series with R*, Springer, New York (2009).
- [3] Crawley M., *Statistics: an introduction using R*, Wiley J., Chichester West Sussex England (2005).
- [4] Erzgräber H., Strozzi F., Zaldívar J. M., Touchette H., Gutiérrez E., and Arrowsmith D. K., *Time series analysis and long range correlations of Nordic spot electricity market data*, Physica A, Volume 387 (2008).
- [5] Exley J., Mehta S., and Smith A., *Mean reversion*, Institute of Actuaries, Finance and Investment Conference, Brussels (2004).
- [6] Heagy J. F., Platt N., and Hammel S. M., *Characterization of on-off intermittency*, Physical Review E, Volume 49 (1994).
- [7] Hull J. C., *Options, Futures, and Other Derivatives*, 8th Edition, Prentice-Hall, Englewood Cliffs, NJ (2012).
- [8] Klüppelberg C., Lecture notes in “*Time Series Analysis*”, Zentrum Mathematik, Technische Universität München, Winter Semester (2015).
- [9] Perelló J., Montero M., Palatella L., Simonsen I., and Masoliver J., *Entropy of the Nordic electricity market: anomalous scaling, spikes, and mean-reversion*, J. Stat. Mech., (2006).
- [10] Shumway R. H., and Stoffer D. S., *Time series analysis and its applications with R examples*, 3rd edition, Springer (2011).
- [11] Simonsen I., *Measuring anti-correlations in the Nordic electricity spot market by wavelets*, Physica A, Volume 322 (2003).

- [12] Simonsen I., *Volatility of power markets*, Physica A 355 (2005).
- [13] Tsay R. S., *Analysis of financial time series*, Wiley-Interscience (2005).
- [14] Weron R. and Przybyłowicz B., *Hurst analysis of electricity price dynamics*, Physica A, Volume 283 (2000).
- [15] Weron R., Bierbrauer M., and Truck S., *Modelling electricity prices: Jump diffusion and regime switching*, Physica A, Volume 336 (2004).
- [16] Weron R., *Energy price risk management*, Physica A, Volume 285 (2000).
- [17] Weron R., Simonsen I., and Wilman P., *Modeling highly volatile and seasonal markets: evidence from the Nord Pool electricity market*, The Application of Econophysics, Springer-Verlag, Tokyo (2004).