NTNU – Trondheim
Norwegian University of
Science and Technology

Extreme Value Analysis & Application of the ACER Method on Electricity Prices

Torgeir Anda

Master of Science in Statistics
Submission date: July 2012
Supervisor: Arvid Næss, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Preface

This thesis concludes my master's degree in statistics (M.Sc.) at the Norwegian University of Science and Technology (NTNU), under the Department of Mathematical Science (IMF) and supervision of Arvid Næss.

By exploring electricity prices in the Nord Pool spot market, we have sought to quantify statistical properties of the distribution of extreme prices.

I would like to give special thanks to my supervisor Arvid Næss, Eirik Mo from Statkraft who has shared valuable insight into the power industry, Kai-Erik Dahlen who has provided code for ACER implementation, and Nord Pool for supplying electricity prices data. I would also like to thank Raimund Maulwurf, Anders Engan, Judith Zehetgruber, Karoline Skogø, Henrik Hemmen and Christian Page

Torgeir Anda
Trondheim, July 2012

Abstract

In this thesis we have explored the very high prices that sometimes occurs in the Nord Pool electricity market Elspot. By applying AR-GARCH time series models, extreme value theory, and ACER estimation techniques, we have sought to estimate the probabilities of threshold exceedances related to electricity prices. Of particular concern was the heavy-tailed Fréchet distribution, which was the asymptotic distribution assumed in the ACER estimation.

We have found that with extreme value theory we are better equipped to deal with the very high quantiles in the time series we have analyzed. We have also described a method that can give an assessment of the probability of exceeding a selected level in the electricity price.

Sammendrag

I denne oppgaven har vi utforsket de meget høye prisene som noen ganger oppstår i Nord Pool kraftmarkedet Elspot. Ved å bruke AR-GARCH tidsrekkemodeller, ekstremverditeori, og ACER estimeringsteknikker, har vi forsøkt å anslå sannsynligheten for terskeloverskridelser knyttet til strømprisen. Av spesiell interesse var den tunghalede Fréchet fordelingen, som var den asymptotiske fordelingen antatt i ACER estimeringen.

Vi har funnet at vi med ekstremverditeori står bedre rustet til å takle de svært høye kvantilene i tidsrekkene vi har analysert. Vi har også beskrevet en metode som kan gi en vurdering av sannsynligheten for å overskride et valgt nivå i strømprisen.

Contents

Preface **1**

Abstract **2**

Sammendrag **3**

1 Introduction **6**

2 The Electricity Market **7**
2.1 Nord Pool 7
2.2 Price Setting at Nord Pool 8
2.3 Extreme prices at Nord Pool 9

3 Data **11**
3.1 The Elspot Series 11
3.2 The Returns Series 13
3.3 Daily Data 17
3.4 Quantile Behavior 21

4 Extreme Value Theory **25**
4.1 Extreme Value Distributions 25
4.2 Peaks Over Threshold 27
4.3 Return Period 28

5 Average Conditional Exceedance Rates **30**
5.1 Cascade of Conditioning Approximations 30
5.2 Empirical Estimation of ACER 32
5.3 Fréchet Fit Optimization 34

6 Modeling Electricity Prices **35**
6.1 In Literature 35
6.2 AR-GARCH model 36

7 Data Analysis **39**
7.1 Analysis of Net Returns 39
7.2 Analysis of Returns 44
7.3 Forecasting Extreme Values 52

8 Returns & Extreme Values **54**
8.1 A Stock Market Comparison 55
8.2 Mean Reversion 55
8.3 Asymmetric Transformation of Extreme Quantiles 60

9 An Alternative Transformation **62**
9.1 A Moving Median 63
9.2 Daily Differences 66

10 ACER Analysis of Differences **69**
10.1 Daily Differences 69
10.2 Hourly Differences 74
10.3 Forecasting With Differences 77

11 Discussion & Concluding Remarks **80**

Bibliography **84**

Chapter 1

Introduction

High electricity prices are always a burden for consumers, as they have no choice but to pay whatever the cost is. In longer periods of high prices, like in winter season 2009/2010 in Trondheim, this often becomes a topic in the media and public debate. It is however not clear to most people the intricacies that lie behind, and that cause high electricity prices.

In this thesis we will begin by giving a rundown of the electricity market in Chapter 2. We will cover some of its history, how it works, and point to some causes that contribute to the very high electricity prices we sometimes observe. After a more qualitative introduction to the electricity market, we will present the data set we have used in Chapter 3, together with some diagnostic plots and some minor remarks.

Motivated by the data we will go through some of the theory in the field of extreme value statistics in Chapter 4 and Chapter 5, and also some ways to model electricity prices in Chapter 6. By applying this theory and one of the models from Chapter 6, we will try to answer some basic questions: (i) How well does extreme value theory apply to describe very high electricity prices? (ii) Can we we use extreme value theory to asses the probability of very high electricity prices?

We begin answering these questions in the data analysis in Chapter 7. Following the data analysis is an exploration of a different methodology that we have experimented with in Chapters 8 through 10. The thesis is rounded off with a discussion and concluding remarks in Chapter 11.

To implement the models and do calculations in this thesis we have used the software R [11], and Matlab [6].

Chapter 2

The Electricity Market

2.1 Nord Pool

Nord Pool was established in 1992 as a consequence of the Norwegian energy act of 1991 that formally paved the way for the deregulation of the electricity sector in Norway [10]. It started out as a norwegian market, but Sweden (1996), Finland (1998) and Denmark (2000) joined in later.

Nord Pool is a commodity market for electricity and can be divided into two parts, *Elbas* and *Elspot*. Both of these markets are for *physical* delivery of power. Elbas is a continuous hour-ahead market, also called the balancing market, and Elspot is a day-ahead market.

There used to be a third market called *Eltermin* which dealt in power derivatives, like forwards (up to three years ahead), futures (up to 8 weeks), options and CFDs. This market is now part of Nasdaq OMX Commodities and is a purely financial market used for hedging and speculation. The different types of contracts listed above uses the Nord Pool Spot Price as their reference price.

The amount of consumed electricity traded on Nord Pool has grown ever since Nord Pool opened, and today about 75% of the total power consumption in the Nordic region is traded on Nord Pool.

2.2 Price Setting at Nord Pool

At Elspot one-hour-long physical power contracts are traded at a minimum unit of 0.1 MWh. At 12 pm each day, the market participants submit to Nord Pool their bid and ask offers for the *next* 24 hours starting at 1 am the next day.

Today there are around 350 buyers and sellers (called members) on Elspot. Most of them trade every day, placing a total of around 2000 orders for power contracts on a daily basis.

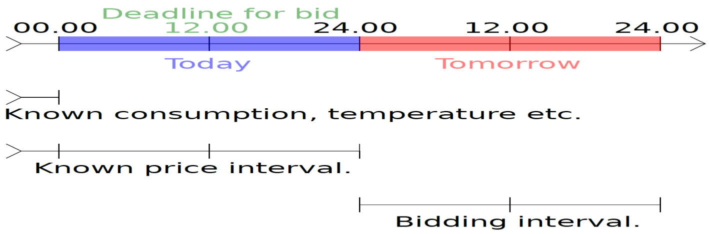

Figure 2.1: *Illustration of the bidding structure at Nord Poll. Source: Kim Stenshorne's master thesis [12].*

Figure 2.1 visualizes the bidding structure and at what times during the day the prices are known. After bids have been placed, supply and demand for each hour is tallied up, and a single price for each hour is found.

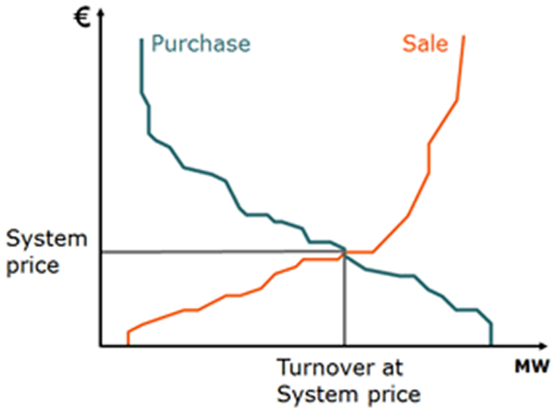

Figure 2.2: *Illustration of the supply and demand curves of spot prices. Source: Nord Pool [9].*

Figure 2.2 illustrates how the price is set after the bids have been placed. This is done individually for every hour the next day.

2.3 Extreme prices at Nord Pool

One of the defining features of the Elspot market is the extreme *volatility*. Even compared to the most volatile commodity markets, none come close to the elspot market. A big part of this volatility is attributed to large short term price changes called *spikes*.

It is hard to say what exactly is causing these spikes. Figure 2.3 attempts to explain part of the reason. When sudden changes in supply (i.e. defect power line) or demand (i.e. it is very cold and more electricity is needed for heating) occurs, sometimes it is necessary to use more expensive sources of energy.

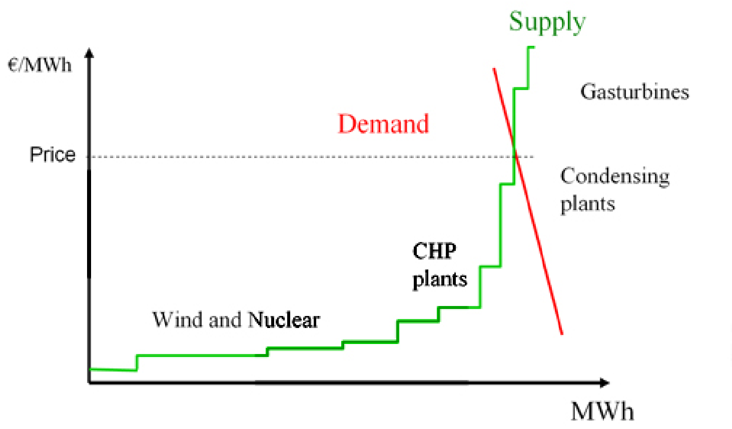

Figure 2.3: *Illustration of different energy sources cost. Source: Wind Energy, The Facts [14].*

Rafal Weron argues in his book *Modeling and Forecasting Electricity Loads and Prices* [13] that this alone does not explain the price spikes. According to him it is the bidding strategies used by the market players that cause the spikes. For many of the market players, electricity is an essential commodity which they are willing to pay almost any price for. The suppliers are aware of this and try to place their own bids accordingly, so as to maximize their profits.

There is a technical ceiling of the Elspot price of 2000 Euros [16], but in reality there is no cap. The highest price we observerd in our data series was 300 Euros.

Chapter 3

Data

Our data are Elspot prices from January 1, 2005 to December 31, 2011. The Elspot market has evolved and changed since it was first established [10], so we thought this to be a reasonable time frame.

The Elspot prices were supplied by Nord Pool and Eirik Mo at Statkraft.

3.1 The Elspot Series

We denote the time series of Elspot prices P_t, where t is hours.

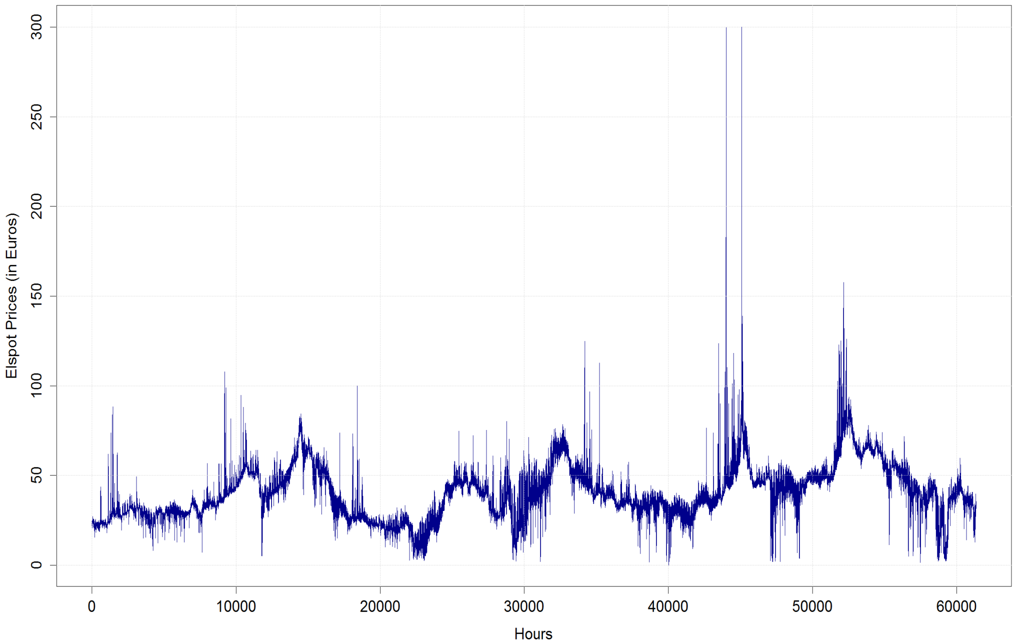

Figure 3.1: *Elspot prices, January 1, 2005 to December 31, 2011.*

From Figure 3.1 we quickly notice the spikes and erratic behavior of the Elspot prices.

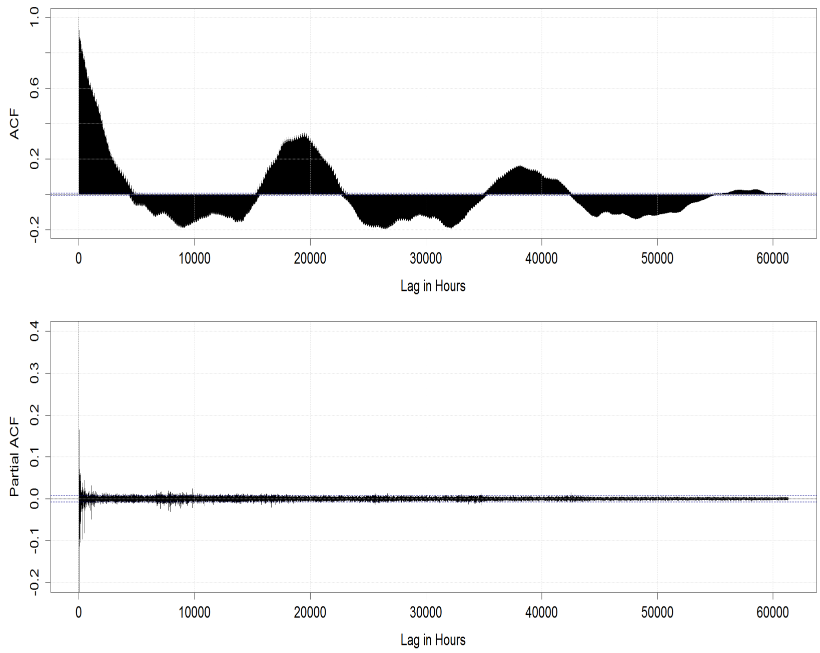

Figure 3.2: *ACF and PACF for Elspot prices with lags for the entire time frame of the data.*

From the PACF and ACF of the Elspot Prices, shown in Figure 3.2, we can see that the long term trends are dominating the ACF.

3.2 The Returns Series

We will call price differences from one hour to the next *returns*. The returns series is formed by differencing P_t

$$R_t = P_t - P_{t-1} \tag{3.1}$$

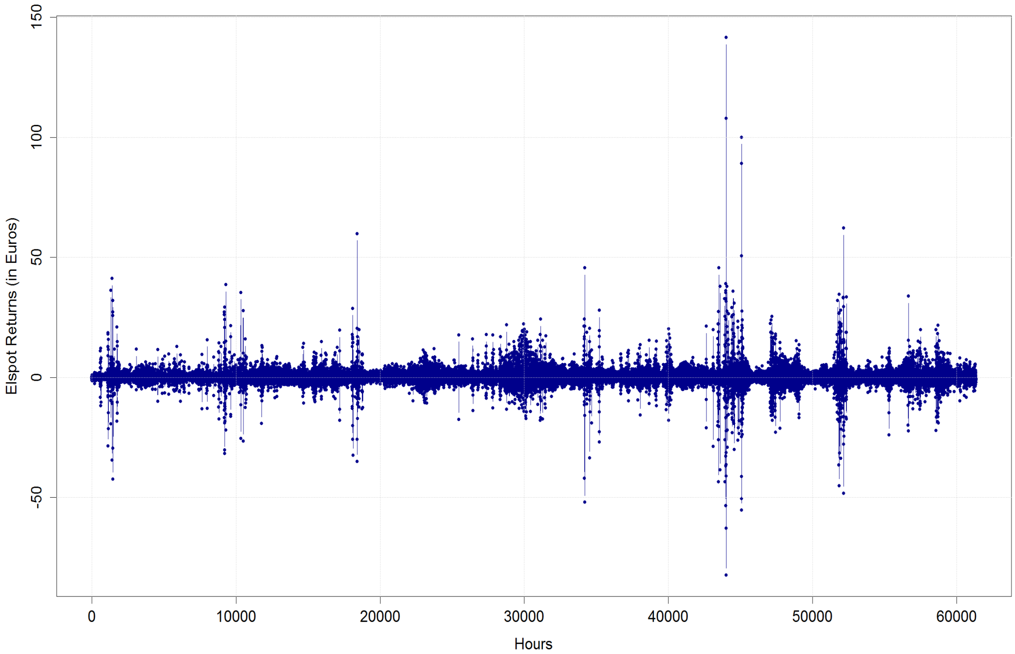

Figure 3.3: *Elspot price changes (returns), January 1, 2005 to December 31, 2011.*

From Figure 3.3 we can see that the return series has some symmetry about the x-axis. There are still spikes in the return series, but the highest value is about half the value of the highest price.

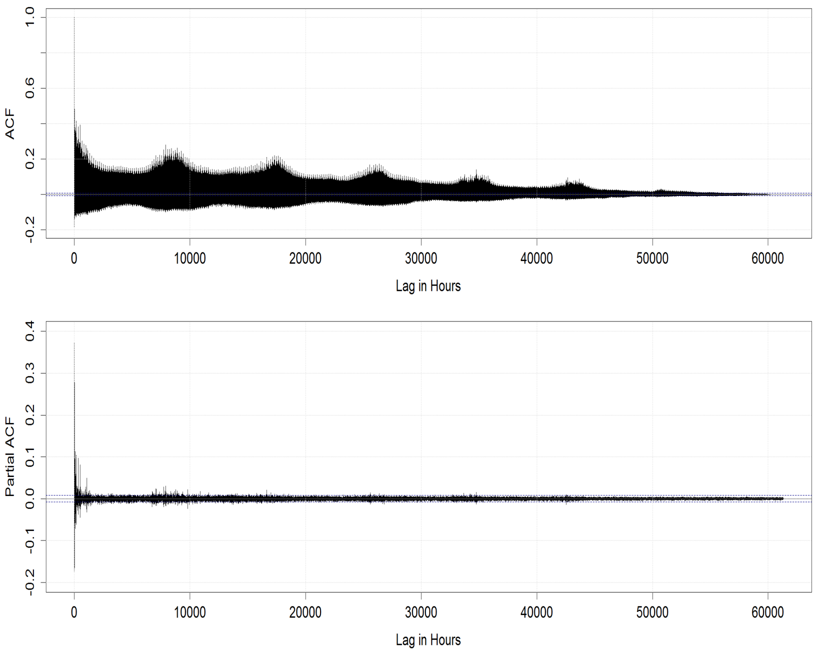

Figure 3.4: *ACF and PACF for Elspot returns with lags for the entire time frame of the data.*

For the returns series in Figure 3.4 we can see more regular behavior in the ACF. There is clearly a yearly pattern in the ACF indicated by the six tops in the ACF (the time series is over seven years, but for last lags there is almost no correlation left).

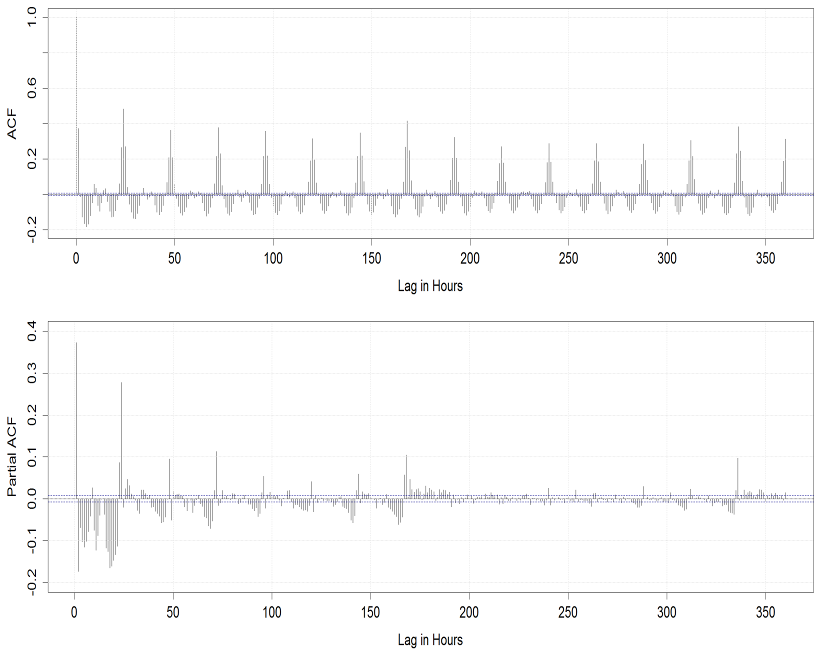

Figure 3.5: *ACF and PACF for Elspot returns with 15 days of lag.*

By taking a look at the ACF with 15 days lag (or 360 hours), we clearly see in Figure 3.5, a daily pattern and a weekly pattern.

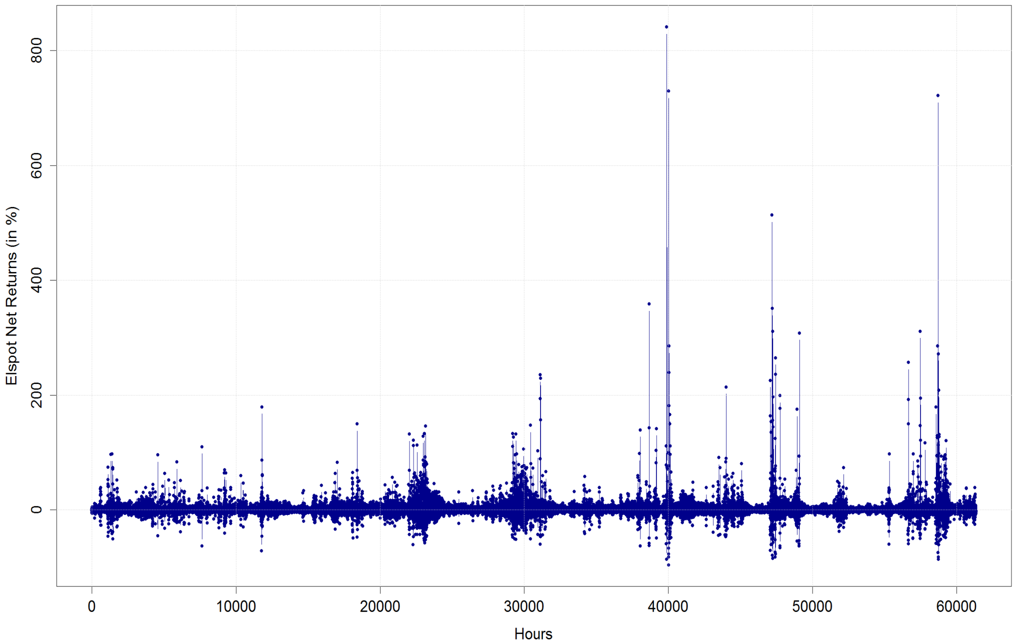

Figure 3.6: *Elspot price changes (%), January 1, 2005 to December 31, 2011.*

We will call relative price changes *net returns*, calculated by

$$N_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{3.2}$$

From Figure 3.6 it looks like the up-spikes have been stretched out compared to the returns series, while the negative relative returns have been compressed. We can imagine the effect of dividing by very low prices, i.e. close to zero.

The ACF and PACF behave in a similar manner as for the returns series.

3.3 Daily Data

The Elspot prices can also be considered as time series of *daily* data. We can create time series of daily data for the Elspot prices by letting the price of a particular hour each day be followed by the price of that same hour the next day.

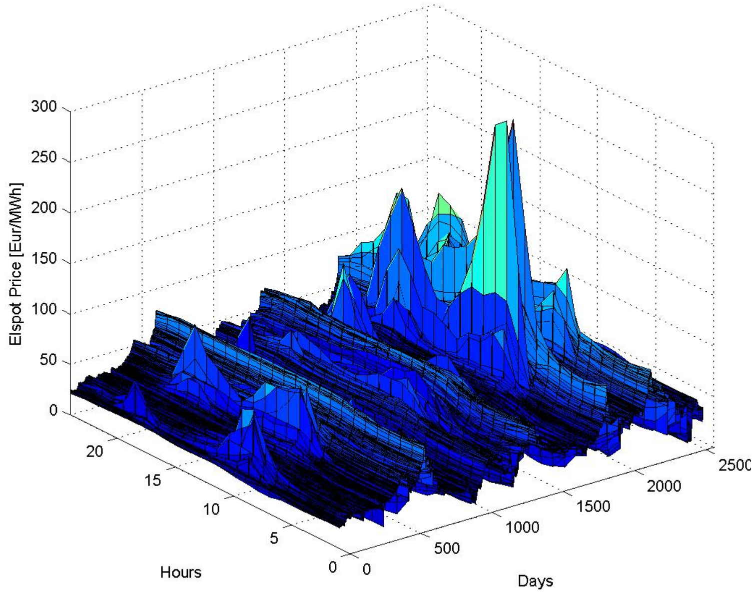

Figure 3.7: *Surface plot of the Elspot price from January 1, 2005 to December 31, 2011*

In this way we can create a total of twenty-four time series for the Elspot prices, one for each hour of the day.

In Figure 3.7 these time series are visualized together, each starting at Saturday, January 1, 2005 (1 Days) and ending Saturday, December 31, 2011 (2556 Days). 1 Hours corresponds to the time-interval 00:00-01:00, and so on. From this figure it is obvious that the daily prices time series, where one hour of the day is selected, behave differently.

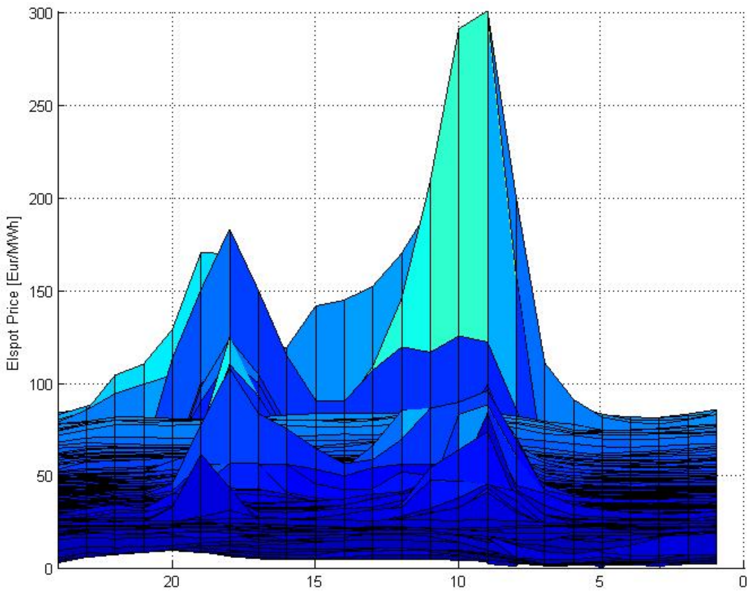

Figure 3.8: *Surface plot of the Elspot price from January 1, 2005 to December 31, 2011*

From Figure 3.8 we can see that the peaks occur at around 08:00-09:00 and 17:00-18:00 hours, and these times also looks to be more volatile. Around 02:00-03:00 hours looks to be the least volatile time. We examine the ACF of the most regular and least regular series, 02:00-03:00 and 08:00-09:00 hours.

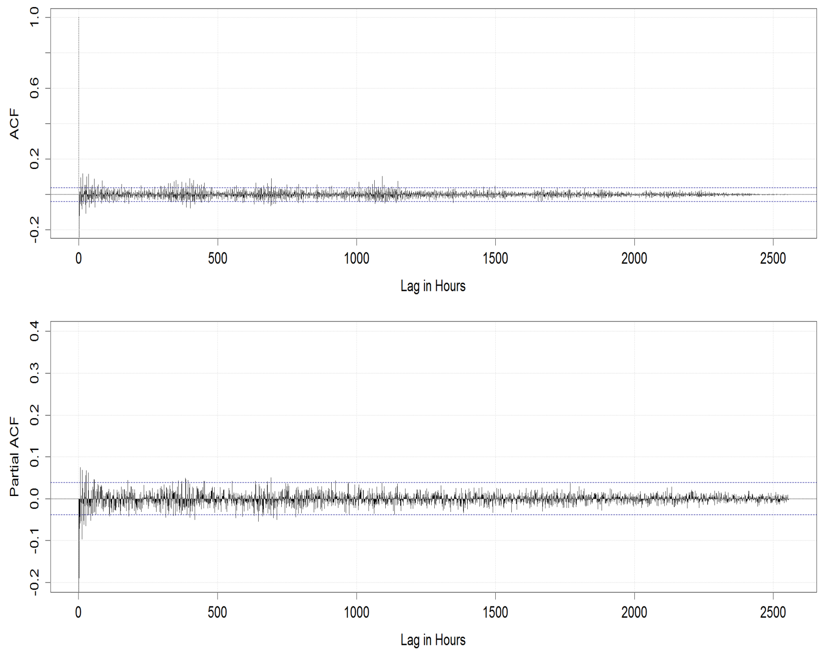

Figure 3.9: *ACF and PACF for 02:00-03:00.*

We see in Figure 3.9 that compared to the returns series, there is hardly any correlation left for hours 02:00-03:00.

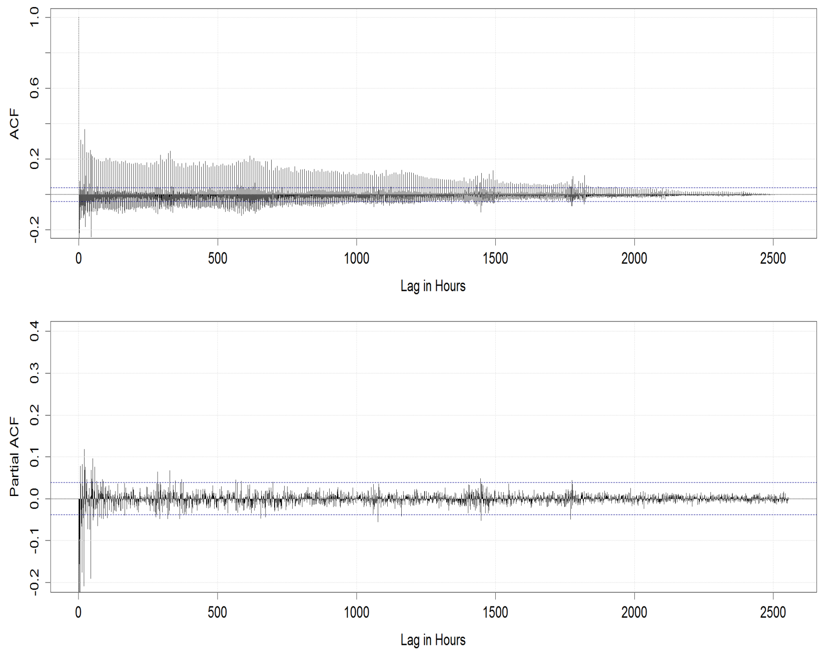

Figure 3.10: *ACF and PACF for 08:00-09:00*

Compared to the ACF for 02:00-03:00 hours, the ACF for 08:00-09:00 hours shows much more correlation. By examining closer in Figure 3.10 we see that the significant correlation comes from the *weekly* lag.

3.4 Quantile Behavior

We take a look at how the normal distribution fits to our Elspot prices, returns, and net returns series.

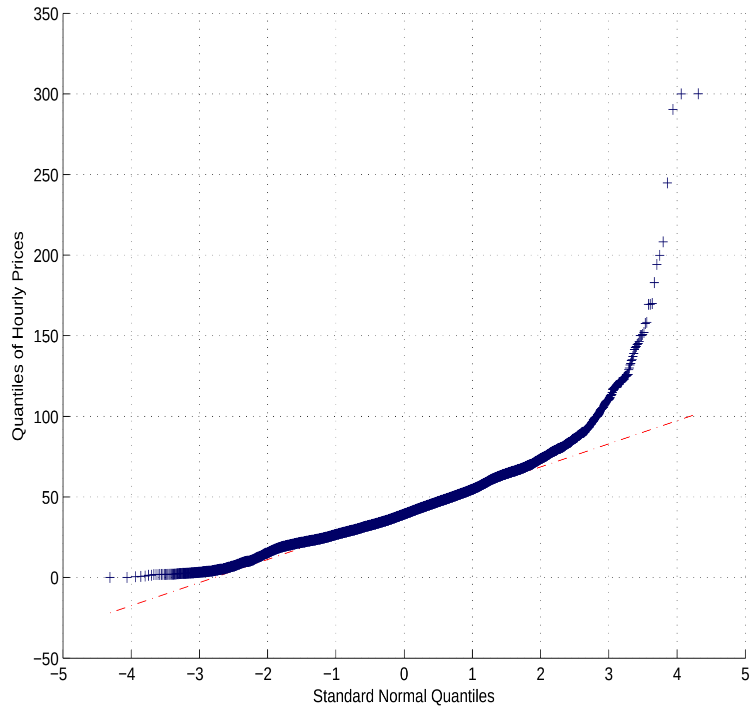

Figure 3.11: *Q-Q plot of the Elspot prices*

For the Elspot prices the normal distribution actually predicts more low prices. This is because the prices cannot be negative, so the data is truncated at zero. For the very high quantiles we can see that the normal distribution predict much lower prices.

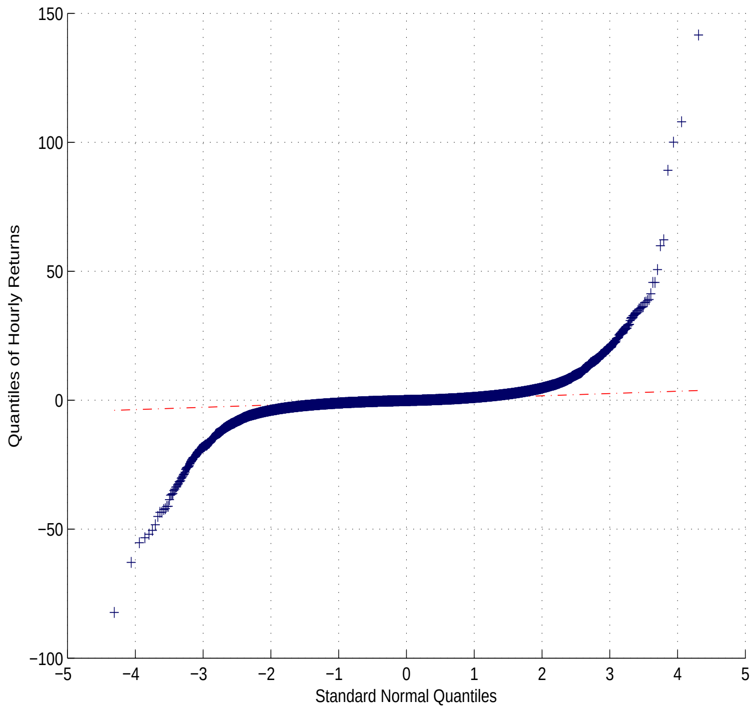

Figure 3.12: *Q-Q plot of the Elspot returns*

As we expected the tail behavior of the returns series deviates significantly from the tail behavior of the normal distribution, both for negative and positive returns.

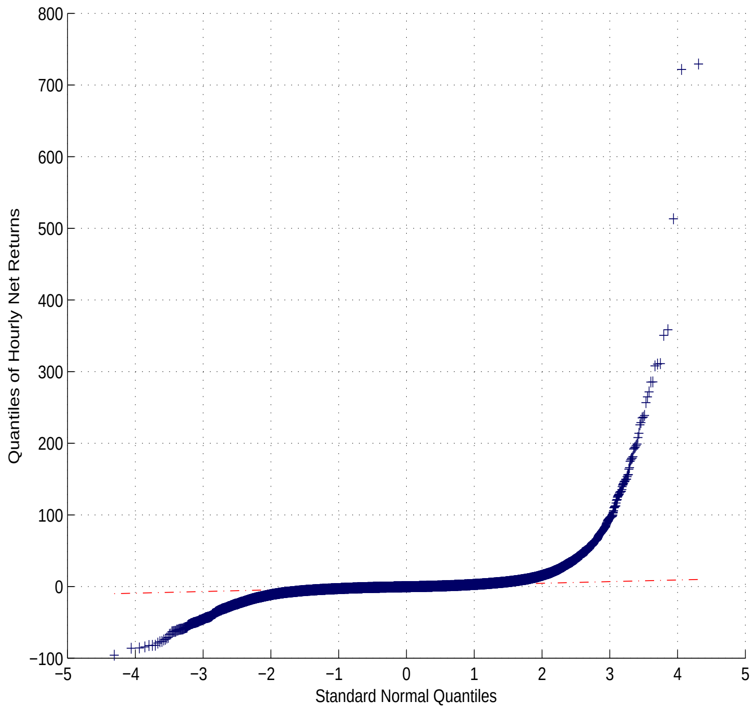

Figure 3.13: *Q-Q plot of the Elspot net returns*

For the very high quantiles it seems that the net returns series has even more extreme behavior.

Chapter 4

Extreme Value Theory

To further explore the extreme prices we have observed in the Elspot market we need to establish a theoretical framework that can be used to analyze the data. We will start off with outlining some basic premises for extreme value analysis as described in *An Introduction to Statistical Modeling of Extreme Values* (2001) [2].

Here extreme values are related to maximum values by considering

$$M_n = \max\{X_1, ..., X_n\}, \tag{4.1}$$

where $X_1, ..., X_n$ is a sequence of independent and identically distributed (iid) random variables having a common distribution function F. The distribution of M_n is then given by

$$\mathrm{P}(M_n \leq \eta) = \mathrm{P}(X_1 \leq \eta) \cdots \mathrm{P}(X_n \leq \eta) = F_M(\eta). \tag{4.2}$$

However F is usually unknown in applications, so we need some other way of finding the distribution of M_n.

4.1 Extreme Value Distributions

We begin with a definition

Definition A distribution G is said to be **max-stable** if, for every $n = 2, 3, ...$, there are constants $\alpha_n > 0$ and β_n such that

$$G^n(\alpha_n \eta + \beta_n) = G(\eta) \tag{4.3}$$

△

Max-stable distributions and extreme value distributions are related in the following way

Theorem 4.1.1. *A distribution is max-stable if, and only if, it is a generalized extreme value distribution.* □

Theorem 4.1.1 is used in the proof, which we will omit, of the following theorem called *Extremal Types Theorem*

Theorem 4.1.2. *If there exist sequences of constants* $\{a_n > 0\}$ *and* $\{b_n\}$ *such that*

$$P\{(M_n - b_n)/a_n \leq \eta\} \to G(\eta) \text{ as } n \to \infty \tag{4.4}$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$\text{I} \ : G(\eta) = \exp\left\{-\exp\left[-\left(\frac{\eta - b}{a}\right)\right]\right\}, \quad -\infty < \eta < \infty, \tag{4.5}$$

$$\text{II} \ : G(\eta) = \begin{cases} 0, & \eta \leq b, \\ \exp\left\{-\left(\frac{\eta-b}{a}\right)^{-\alpha}\right\}, & \eta > b, \end{cases} \tag{4.6}$$

$$\text{III} \ : G(\eta) = \begin{cases} \exp\left\{-\left[-\left(\frac{\eta-b}{a}\right)^{\alpha}\right]\right\}, & \eta < b, \\ 1, & \eta \geq b, \end{cases} \tag{4.7}$$

for parameters $a > 0, b$ *and, in this case of families II and III,* $\alpha > 0$. □

These three families of distributions are called Gumbel, Fréchet and Weibull. They are each special cases of the generalized extreme value (GEV) family. We restate Theorem 4.1.2 using the generalized form

Theorem 4.1.3. *If there exist sequences of constants* $\{a_n > 0\}$ *and* $\{b_n\}$ *such that*

$$P\{(M_n - b_n)/a_n \leq \eta\} \to G(\eta) \text{ as } n \to \infty, \tag{4.8}$$

for a non-degenerate distribution function G, then G is a member of the GEV family

$$G(\eta) = \exp\left\{-\left[1 + \xi\left(\frac{\eta - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{4.9}$$

defined on $\eta : 1 + \xi(x - \mu)/\sigma > 0$, *where* $-\infty < \mu < \infty, sigma > 0$ *and* $-\infty < \xi < \infty$. □

In plainer words Theorem 4.1.3 says that if there exist a distribution function for $\frac{M_n - b_n}{a_n}$, then it must be on the form of Equation 4.9. The apparent problem with the constants b_n and a_n can be solved by assuming

$$\mathrm{P}\left\{(M_n - b_n)/a_n \leq \eta\right\} \approx G(x) \tag{4.10}$$

for large enough n. Equivalently

$$\mathrm{P}\{M_n \leq \eta\} \approx G\left\{(\eta - b_n)/a_n\right\} \tag{4.11}$$
$$= G^*(\eta), \tag{4.12}$$

where G* is another member of the GEV family. In other words, if Theorem 4.1.3 enables approximation of the distribution of M_n^* by a member of the GEV family for large n, the distribution of M_n itself can also be approximated by a different member of the same family. Since the parameters of the distribution have to be estimated anyway, it is irrelevant in practice that the parameters of the distribution G are different from those of G^*.

Thus we can partition a data series into blocks, and use the maximum of those blocks to estimate an extreme value distribution for M_n.

4.2 Peaks Over Threshold

We explained in Section 4.1 how we can construct block maximums to estimate M_n. A problem with this is that the extreme values might not be evenly spread throughout the data series, so we might end up discarding a lot of extreme values if one or more blocks contains several extreme values.

The following theorem enables us to pick out extreme values in another way

Theorem 4.2.1. *Let $X_1, X_2, ...$ be a sequence of independent random variables with common distribution function F, and let*

$$M_n = \max\{X_1, ..., X_n\}. \tag{4.13}$$

Denote an arbitrary term in the X_i sequence by X, and suppose that F satisfies Theorem 4.1.2, so that for large n,

$$Pr\{M_n \leq \eta\} \approx G(\eta), \tag{4.14}$$

where

$$G(\eta) = \exp\left\{ -\left[1 + \xi\left(\frac{\eta - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{4.15}$$

for some $\mu, \sigma > 0$ and ξ. Then, for large enough u, the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{4.16}$$

defined on $\{y : y > 0$ and $(1 + \xi y/\tilde{\sigma}) > 0\}$, where

$$\tilde{\sigma} = \sigma + \xi(\eta - \mu). \tag{4.17}$$

□

Theorem 4.2.1 states that if F satisfies Theorem 4.1.2, so that for large n M_n follows a generalized extreme value distribution, then for a high enough threshold u, the threshold exceedances will follow a *generalized Pareto distribution.*

This result becomes very useful because it enables us to filter out extreme values in a new way. Instead of just using block maxima, we can select a threshold u and use all the data values that exceeds this threshold.

Dependence in Threshold Exceedances

A problem that may arise when using POT methods instead of block maxima, is that the threshold exceedances may be clumped together. This would indicate, in most cases, that the independence assumption from Equation 4.1 has been violated.

A common way to deal with this is *declustering* where in the simplest case an extreme value would be disregarded if it was in close proximity to another extreme value.

4.3 Return Period

[3]

The return period of a level η for a random variable X is defined as

$$R = \frac{1}{P(X > \eta)} = \frac{1}{1 - F_X(\eta)}. \tag{4.18}$$

This means that the return rate R for η is the mean number of trials that must be done for X to exceed η.

When modeling POT the return period for a level $\eta_R = u + y$, where u is the threshold, is given by

$$R = \frac{1}{\lambda P(X > \eta_R)} = \frac{1}{\lambda P(Y > y)}. \tag{4.19}$$

Here λ is the mean crossing rate of the threshold per block (i.e. per year, month etc.), or the average proportion of observations that fall over the threshold. From (4.19) it follows that

$$P(Y \leq y) = 1 - \frac{1}{\lambda R}, \tag{4.20}$$

and since the distribution of Y is known, we have from ... that

$$\eta_R = u - \frac{\tilde{\sigma}}{\xi}\left(1 - (\lambda R)^{\xi}\right), \tag{4.21}$$

for $\xi \neq 0$, and

$$\eta_R = u + \tilde{\sigma}\log(\lambda R), \tag{4.22}$$

for $\xi = 0$.

Confidence intervals for the return level η_R is computed using the delta method, that is assuming that the maximum likelihood estimator is multinormal distributed with expectation equal to the real parameter value and variance covariance matrix V. The variance of the return level η_R can tehn be estimated by the delta method as

$$Var(\eta_R) \approx \nabla\eta_R^T V \nabla\eta_R, \tag{4.23}$$

where V is the variance-covariance matrix for the estimated parameters $(\hat{\lambda}, \hat{\sigma}, \hat{\xi})$, and

$$\nabla\eta_R = \left[\frac{\partial\eta_R}{\partial\lambda}, \frac{\partial\eta_R}{\partial\sigma}, \frac{\partial\eta_R}{\partial\xi}\right]^T. \tag{4.24}$$

For the ACER method the return levels are estimated by inverting ... with the values for a, b, c and q, found by the least square routine, for the exceedance rate of interest.

Chapter 5

Average Conditional Exceedance Rates

The concept of average conditional exceedance rates (ACER) is a relatively new method for extreme value estimation. The method is developed by Arvid Næss, and is among others explained in the preprint *Estimation of Extreme Values of Time Series with Heavy Tails* (2010) [8].

The main differences between ACER and POT methods are that ACER gives an exact *empirical* distribution without making assumptions about independence in the data.

5.1 Cascade of Conditioning Approximations

We let $0 \leq t_1 < ... < t_N \leq T$ denote the points in time for the observerd data values of $X(t)$, and let $X_k = X(t_k), k = 1, ..., N$. We use the notation from Equation 4.1 and denote $\mathrm{P}(M_n \leq \eta) = \mathrm{P}(\eta)$. $\mathrm{P}(\eta)$ is then given exactly by

$$\begin{aligned} \mathrm{P}(\eta) &= \mathrm{P}\{X_1 \leq \eta, ..., X_N \leq \eta\} \\ &= \mathrm{P}\{X_N \leq \eta | X_1 \leq \eta, ..., X_{N-1} \leq \eta\} \cdot \mathrm{P}\{X_1 \leq \eta, ..., X_{N-1} \leq \eta\} \\ &= \prod_{j=2}^{N} \mathrm{P}\{X_j \leq \eta | X_1 \leq \eta, ..., X_{j-1} \leq \eta\} \cdot \mathrm{P}(X_1 > \eta). \end{aligned} \tag{5.1}$$

If we assume that the Xs are iid (∼Poisson appr.), then with $\alpha_{1j}(\eta) = \mathrm{P}\{X_j > \eta\}$ we have

$$\mathrm{P}(\eta) \approx \prod_{j=1}^{N} \mathrm{P}(X_j \leq \eta) = \prod_{j=1}^{N}(1 - \alpha_{1j}(\eta))$$
$$\approx \mathrm{P}_1(\eta) = \exp\{-\sum_{j=1}^{N} \alpha_{1j}(\eta)\}. \quad (5.2)$$

If we instead of assuming independence condition on one previous value we have

$$\mathrm{P}\{X_j \leq \eta | X_1 \leq \eta, ..., X_{j-1} \leq \eta\} \approx \mathrm{P}\{X_j \leq \eta | X_{j-1} \leq \eta\}, \quad (5.3)$$

which leads to the approximation

$$\mathrm{P}(\eta) \approx \mathrm{P}_2(\eta) = \exp\left\{-\sum_{j=2}^{N} \alpha_{2j}(\eta) - \alpha_{11}(\eta)\right\}, \quad (5.4)$$

where

$$\alpha_{2j} = \mathrm{P}\{X_j > \eta | X_{j-1} \leq \eta\}. \quad (5.5)$$

We can continue with conditioning on more values

$$\mathrm{P}(\eta) \approx \mathrm{P}_3(\eta) = \exp\left\{-\sum_{j=3}^{N} \alpha 3j(\eta) - \alpha 22(\eta) - \alpha 11(\eta)\right\}, \quad (5.6)$$

where

$$\alpha_{3j}(\eta) = \mathrm{P}\{X_j > \eta | X_{j-1} \leq \eta, X_{j-2} \leq \eta\}, \quad (5.7)$$

and

$$\mathrm{P}(\eta) \approx \mathrm{P}_4(\eta) = \exp\left\{-\sum_{j=4}^{N} \alpha 4j(\eta) - \alpha 33(\eta) - \alpha 22(\eta) - \alpha 11(\eta)\right\}, \quad (5.8)$$

where

$$\alpha_{4j}(\eta) = \mathrm{P}\{X_j > \eta | X_{j-1} \leq \eta, X_{j-2} \leq \eta, X_{j-3} \leq \eta\}. \quad (5.9)$$

This process of conditioning on more and more previous values can be continued until there are no more values, but in particular, if $N >> k$, where k is the number of values we condition on, we can approximate

$$\mathrm{P}_k(\eta) \approx \exp\left\{-\sum_{j=k}^{N} \alpha_{kj}(\eta)\right\},\ k = 1, 2, ... \tag{5.10}$$

By using the approximation in Equation 5.10 we introduce the concept of ACER in the following way

$$\varepsilon_k(\eta) = \frac{1}{N-k+1}\sum_{j=k}^{N} \alpha_{kj}(\eta),\ k = 1, 2, ... \tag{5.11}$$

5.2 Empirical Estimation of ACER

The following random functions are defined

$$A_{kj}(\eta) = \mathbf{1}\{X_j > \eta, X_{j-1} \leq \eta, ..., X_{j-k+1} \leq \eta\},\ j = k, ..., N,\ k = 2, 3, \tag{5.12}$$

and

$$B_{kj}(\eta) = \mathbf{1}\{X_{j-1} \leq \eta, ..., X_{j-k+1} \leq \eta\},\ j = k, ..., N,\ k = 2, ..., \tag{5.13}$$

where $\mathbf{1}\{A\}$ denotes the indicator function of some event A. Then

$$\alpha_{kj}(\eta) = \frac{E[A_{kj}(\eta)]}{E[B_{kj}(\eta)]},\ j = k, ..., N,\ k = 2, ..., \tag{5.14}$$

where $E[\cdot]$ denotes the expectation operator. Assuming an ergodic process, then obviously $\varepsilon_k(\eta) = \alpha_{kk}(\eta) = ... = \alpha_{kN}(\eta)$, and it may be assumed that

$$\varepsilon_k(\eta) = \lim_{N\to\infty} \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{\sum_{j=k}^{N} B_{kj}(\eta)} \tag{5.15}$$

Clearly, $\lim_{\eta\to\infty} \sum_{j=k}^{N} B_{kj}(\eta) = N-k+1 \approx N$. Hence, $\lim_{\eta\to\infty} \tilde{\varepsilon}_k(\eta)/\varepsilon_k(\eta) = 1$ where

$$\tilde{\varepsilon}_k(\eta) = \lim_{N\to\infty} \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{N-k+1} \tag{5.16}$$

The advantage of this modified ACER function $\tilde{\varepsilon}_k(\eta)$ is that it is easier to use for non-stationary or long-term statistics.

The sample estimate of $\tilde{\varepsilon}_k(\eta)$ is

$$\hat{\varepsilon}_k(\eta) = \frac{1}{R}\sum_{r=1}^{R}\hat{\tilde{\varepsilon}}_k^{(r)}(\eta), \tag{5.17}$$

where

$$\hat{\varepsilon}_k^{(r)}(\eta) = \lim_{N\to\infty}\frac{\sum_{j=k}^{N} A_{kj}^{(r)}(\eta)}{N-k+1}. \tag{5.18}$$

Estimating Confidence Intervals

The sample standard deviation is estimated by the standard formula

$$\hat{s}_k(\eta)^2 = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\varepsilon}_k^{(r)}(\eta) - \hat{\varepsilon}_k(\eta)\right)^2 \tag{5.19}$$

We can use the sample standard deviation to create a good approximation of the 95% CI for $\tilde{\varepsilon}_k(\eta)$ where

$$\mathrm{CI}^{\pm}(\eta) = \hat{\varepsilon} \pm 1.96\hat{s}_k(\eta)/\sqrt{R} \tag{5.20}$$

Fitting Asymptotic Distributions

From Equations 4.5-4.7 we recall that there are three families of extreme value distributions, the Gumbel, Fréchet and Weibull. Based on prior knowledge that Elspot prices seem to have so-called fat-tailed behavior, we will focus on the Fréchet case. By fitting empirical ACER functions to asymptotic distributions, we can use the parametric form to make predictions of extreme values.

If the data are independent this can be expressed as

$$\varepsilon_1(\eta) \approx \left[1 + \xi\left(a(\eta-b)\right)\right]^{-\frac{1}{\xi}}, \quad \eta_{\geq}\eta_0 \tag{5.21}$$

with the first ACER function where a corresponds to $\frac{1}{\sigma}$ and b corresponds to μ in Equation 4.9. This relies on the assumption that sampled data can be used as a basis for prediction.

Since we in practice never have infinite data a form that can capture sub-asymptotic behavior is desirable. Without going into details we will assume that $\varepsilon_k(\eta)$ can be approximated by

$$\varepsilon_k(\eta) \approx q_k(\eta)\left[1 + \xi_k\left(a_k(\eta - b_k)^{c_k}\right)\right]^{-\frac{1}{\xi_k}}, \quad \eta_{\geq}\eta_1 \tag{5.22}$$

where the function $q_k(\eta)$ is weakly varying compared with the function $a_k(\eta - b_k)^{c_k}$. By assuming that $q_k(\eta)$ varies sufficiently slow in the tail region we may replace it with a constant q. We finally write

$$\varepsilon(\eta) \approx q\left[1 + \tilde{a}(\eta - b)^c\right]^{-\gamma}, \quad \eta_{\geq}\eta_1, \tag{5.23}$$

where $\gamma = \frac{1}{\xi}, \tilde{a} = a\xi$.

5.3 Fréchet Fit Optimization

To fit our estimated ACER functions to the parametric form in Equation 5.23 we need an optimization routine.

We define the mean square error function as

$$F(\tilde{a}, b, c, q, \gamma) = \sum_{j=1}^{N} w_j \left|\log \hat{\varepsilon}(\eta_j) - \log q + \gamma\left[1 + \tilde{a}(\eta_j - b)^c\right]\right|^2 \tag{5.24}$$

where $w_j = (\log CI^+(\eta_j) - \log CI^-(\eta_j))^{-2}$ is a weight factor that emphasizes less extreme data points. However this weighting is a matter of preference and application, and can be done in other ways.

To minimize the mean square error function, and to find estimates for the five parameters $\tilde{a}, b, c, q, \gamma$, the Levenberg-Marquardt least squares optimization method is well suited. By fixing $\tilde{a}, b, c$ we find optimal values for

$$\gamma^*(\tilde{a}, b, c) = -\frac{\sum_{j=1}^{N} w_j(x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{N} w_j(x_j - \bar{x})^2} \tag{5.25}$$

and

$$\log q^*(\tilde{a}, b, c) = \tilde{y} + \gamma^*(\tilde{a}, b, c)\bar{x} \tag{5.26}$$

where $y_j = \log \hat{\varepsilon}(\eta_j)$ and $x_j = 1 + \tilde{a}(\eta_j - b)^c$. The Levenberg-Marquardt method is then used on the function $\tilde{F}(\tilde{a}, b, c) = F(\tilde{a}, b, c, q^*(\tilde{a}, b, c), \gamma^*(\tilde{a}, b, c))$ to find the optimal values of $\tilde{a}^*, b^*$ and c^*, and then Equations 5.25 and 5.26 are used to calculate γ^* and q^*.

Chapter 6

Modeling Electricity Prices

In Chapters 4 and 5 we set up some tools for analyzing extreme values. But before we can apply these tools we need to consider the data. As shown in Figure 3.1, the plot of Elspot prices, the data are hardly *stationary*. Stationary means, simply put, that there are no trends or dependence in the data. Ideally we want to transform and model the data in such a way that we end up with *residuals* that are independent and identically distributed to fulfill the assumptions of Equation 4.1. In practice these assumptions rarely hold. We do note that the dependence assumption is not necessary for the ACER method, but will always make things easier.

6.1 In Literature

Our first idea was to use vector auto-regression (VAR) as described in Richard Harris & Robert Sollis' *Applied Time Series Modeling and Forecasting* (2003) [5]. VAR is a statistical model used to capture the linear interdependencies among multiple time series. VAR models generalize the univariate autoregression (AR) models. All the variables in a VAR are treated symmetrically; each variable has an equation explaining its evolution based on its own lags and the lags of all the other variables in the model.

In the implementation of this model we ended up with a huge number of parameters and were unable to limit this with the R functions we used.

Another model we studied was the model Kim Stenshorne used in his master's thesis [12] where a mixed model was proposed. Prices, or rather returns were divided into 'balanced' and 'unbalanced' returns. Ordinary least squares regression was then used to capture the 24-hour correlation between hours, for one day. A generalized least squares regression was then applied to capture between days correlation. This model was fairly successful in predicting for the 'balanced' returns, but did not attempt to make predictions for the 'unbalanced' returns, or extreme prices.

We wanted to incorporate extreme value theory into describing the 'unbalanced' returns, but failed to incorporate all of the rather substantial amount of code included in Kim Stenshore's thesis.

In Rafael Weron's book *Modeling and Forecasting Electricity Loads and Prices* (2006) [13], several models are described. We list some of them here:
Statistical models - *ARMA-type models*, *Time Series with Exogenous Variables*, *Autoregressive GARCH Models*, *Regime Switching Models.*
Quantitative Models - *Jump Diffusion Models*

In talks with Statkraft we learned that they use powerful *fundemental* models that can actually predict, to some degree, even extreme Elspot prices. To accomplish this Statkraft benefits from having an incredible amount of available data (detailed information on power lines, consumption, etc) to use as exogenous variables.

After these meetings we decided that our focus should be on the *stochastic* behavior of the extremes. Fundamental models such as the one Statkraft uses are very powerful for short term and even mid term predictions. So powerful in fact, that they explain the stochastic phenomenon we are trying to study.

For our area of interest we found it best to use pure time series modeling of the data, and consider all the possible exogenous variables as part of an underlying stochastic process.

6.2 AR-GARCH model

In Alexander J. McNeil and Rüdiger Frey's article *Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach* (2000) [7], and Hans N.E. Byström's article *Extreme value theory and extremely large electricity price changes* (2005) [1] it is proposed to use an AR-GARCH model to model returns, and then apply extreme value theory to a series of standardized residuals. We will follow the approach of Byström.

Because of the daily and weekly correlation patterns observed in the returns series we will use an AR filter with lags at $t-1, t-24$ and $t-168$. We could probably get a better fit by including more lags, but we want a simple and intuitive model

$$r_t = a_0 + a_1 r_{t-1} + a_2 r_{t-24} + a_3 r_{t-168} + \epsilon_t \tag{6.1}$$

To account for varying volatility in the time series, a GARCH model is proposed

$$\sigma_t^2 = \phi_0 + \phi_1 \epsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2 \tag{6.2}$$

Again a simple model is chosen with just the first lag for the errors and standard deviations. Making more sophisticated GARCH models could be an interesting study, but not something we will focus on here.

After an AR-GARCH model is fitted to the returns wilth either normal or student's t distributed innovations, we can reverse the process by first picking out quantiles from a fitted extreme value distribution, then scale it with σ_t and finally add to the AR trend.

$$\alpha_{t,p} = a_0 + a_1 r_{t-1} + a_2 r_{t-24} + a_3 r_{t-168} + \sigma_t \alpha_p \tag{6.3}$$

To fit an extreme value distribution, a POT method is applied to the standardized residuals.

POT

The peak over threshold method is an application of Theorem 4.2.1. We use it to estimate α_p.

The values of the standardized residuals series that are over the threshold u follow the excess distribution $F_u(y)$ given by

$$\begin{aligned} F_u(y) &= \mathrm{P}(R - u \leq y R > u) \\ &= \frac{F_R(u+y) - F_R(u)}{1 - F_R(u)}, \quad 0 \leq y \leq R_F - u \end{aligned} \tag{6.4}$$

Equation 4.16 with $\tilde{\sigma}$ equal to α

$$G_{\xi,\alpha}(y) = [1 - (1 + \frac{\xi y}{\alpha})]^{-1/\xi}, \quad \text{if } \xi \not0 \tag{6.5}$$

$$G_{\xi,\alpha}(y) = 1 - e^{-y/\alpha}, \quad \text{if } \xi = 0. \tag{6.6}$$

$$F_R(u+y) = (1 - F_R(u))F_u(y) + F_R(u). \tag{6.7}$$

We write $F_R(u)$ as $(n - N_u)/n$ where n is the total number of observations and N_u is the number of observations above the threshold.

$$F_R(x) = 1 - \frac{N_u}{n}(1 + \frac{\xi}{\alpha}(x-u))^{-1/\xi}. \tag{6.8}$$

and α given by

$$\alpha_p = u + \frac{\alpha}{\xi}\left((\frac{n}{N_u}p)^{-\xi} - 1\right). \tag{6.9}$$

Chapter 7

Data Analysis

We began by analyzing the Elspot net returns series as Byström did in his article [1], but now with newer data. Byström used Elspot prices from January 1, 1996 to October 1, 2000, whereas we have used data from January 1, 2005 to December 31, 2011.

To implement the AR-GARCH model we have used the rugarch library [4] in R. This library allows you model the AR and GARCH part at the same time, and also lets you fix parameters to predefined values.

To model the threshold exceedances we have used the fExtremes library [15] in R. For the ACER implementation we have used Matlab [6] , with help from Kai Erik Dahlen [3].

7.1 Analysis of Net Returns

We repeat the AR-GARCH model we have used for convenience

$$r_t = a_0 + a_1 r_{t-1} + a_2 r_{t-24} + a_3 r_{t-168} + \epsilon_t \quad (7.1)$$

$$\sigma_t^2 = \phi_0 + \phi_1 \epsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2 \quad (7.2)$$

Here r_t corresponds to N_t, the net returns.

	Normal	Student's t
AR-GARCH parameters		
a_0	-0.401 (0.206)	-0.667 (0.0404)
a_1	0.182 (0.00427)	0.168 (0.00321)
a_2	0.294 (0.00348)	0.289 (0.00375)
a_3	0.178 (0.0039)	0.202 (0.0337)
ϕ_0	0.386 (0.0237)	1.24 (0.0337)
ϕ_1	0.249 (0.00552)	0.671 (0.011)
ϕ_2	0.750 (0.00715)	0.328 (0.00624)
v		3.08 (0.0243)
Standardized residuals statistics		
Mean (%)	0.0153	0.02
Standard deviation (%)	1.12	1.26
Skewness	1.68	1.97
Excess kurtosis	38.3	48.6
Q(10)	1519	1484
Q(20)	1810	1762
$Q^2(10)$	35.72	888.9
$Q^2(20)$	62.09	927.2
GPD parameters with POT		
ξ	0.312 (0.0237)	
α	0.884 (0.0259)	
u	0.05	

Table 7.1: *AR-GARCH parameters, statistics on the standardized residuals, as well as GPD parameters for the net returns series*

In Table 7.1 all the parameters from Equations 7.1 and 7.2 are fitted with both normally distributed and student's t distributed innovations. Byström used a 5.5% threshold for his POT analysis, and we have used a similar threshold of 5%, denoted u in Table 7.1. To estimate the POT parameters ξ and α we used the standardized residuals from the AR-GARCH model with normally distributed innovations. The standardized residuals are calculated by dividing the residuals from the auto regressive (AR) filter by the time dependent standard deviations from the GARCH modeling.

In Byström's article he compared how well a normal distribution, a student's t distribution and a generalized Pareto distribution would describe the top

5.5% of the standardized residuals calculated from the AR-GARCH model. He did this by calculating the expected number of exceedances for different quantiles, and then seeing how the empirical exceedances for the different distributions would compare.

Probability	Expected	AR-GARCH	AR-GARCH-t	Conditional GPD
0.95	3068	2812	1585	3067
0.99	614	1409	409	632
0.995	307	1150	246	313
0.999	62	789	75	59
0.9995	31	696	37	30
0.9999	7	514	6	4

Table 7.2: *Empirical exceedances for normal, student's t and GPD distributions compared to the theoretically expected number of exceedances.*

We have repeated this analysis and presented the results in Table 7.2. The numbers are very much in line with Byström's results, demonstrating that the extreme value distribution describes exceedances in the tail of the standardized residuals better. To expand up Byström analysis we wanted to see how ACER would predict exceedances for the same quantiles.

ACER Modeling

To align ourselves with the time dependence structure used in the AR-GARCH model, we study the ACER functions $k = 1, k = 2, k = 25$ and $k = 169$ corresponding to the lags in the AR-filter in Equation 7.1. We wanted the ACER predictions to be comparable to the POT predictions, so we started with the same threshold as for the POT estimation, in the ACER estimation. That is, 5%, or $\eta_1 = 1.56$ for the standardized residuals.

Probability	Expected	$ACER_1$	$ACER_{25}$	$ACER_{169}$
0.95	3068	3103	976	92
0.99	614	610	477	90
0.995	307	310	263	84
0.999	62	63	57	44
0.9995	31	30	29	23
0.9999	7	9	10	6

Table 7.3: *AR-GARCH parameters, statistics on the standardized residuals, as well as GPD parameters*

The results presented in Table 7.3 were a bit surprising. The empirical exceedances based on the ACER fit predictions were very close the to expected number of exceedances for the first ACER function. But the exceedances for $ACER_{25}$ and $ACER_{169}$ were way off except for the most extreme quantiles.

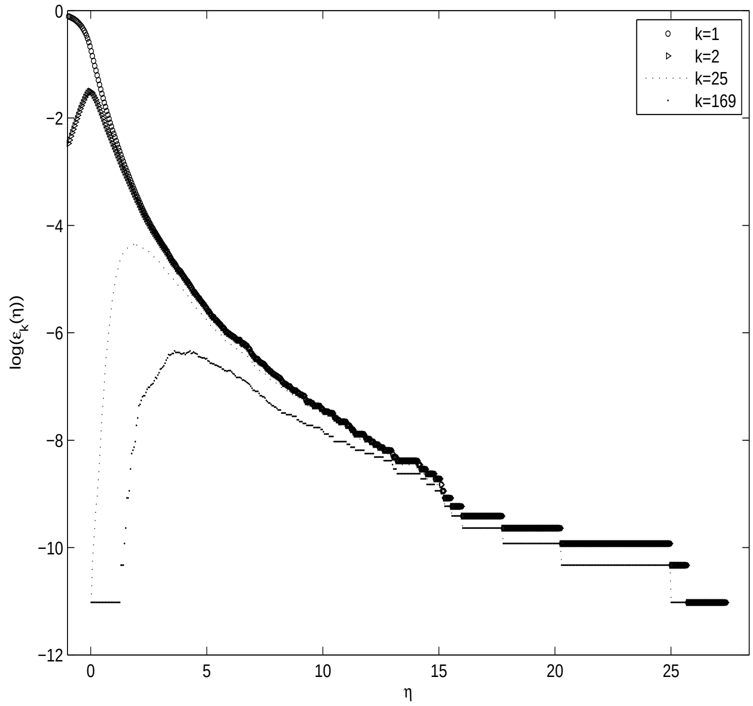

Figure 7.1: *A selection of ACER functions for the standardized residuals from the AR-GARCH fit of net returns.*

By plotting the ACER functions, shown in Figure 7.1, we quickly realized some obvious shortcomings. Our threshold selection of $\eta_1 = 1.56$ looked to be ok for the first ACER function, but for ACER_{169} the tail doesn't start until atleast $\eta_1 = 5$. We therefore needed to choose the tail marker η_1 in a different way.

We will explore this further in the next section, but first we need to explain another issue in the ACER implementation.

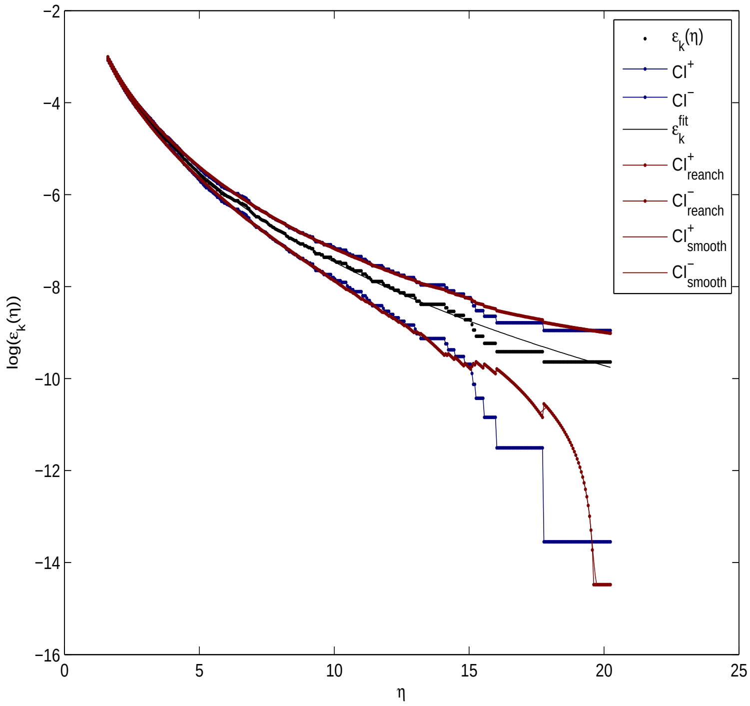

Figure 7.2: *$ACER_1$ with fit for the standardized residuals from the AR-GARCH fit of net returns.*

In Figure 7.2 we show the empirical ACER_1 function together with the parametric fit $q\,[1 + \tilde{a}(\eta - b)^c]^{-\gamma}$ from Equation 5.23. Here we can see that the data are cut at the tail marker $\eta_1 = 1.56$, but if we look closely and compare with Figure 7.1 we notice that the most extreme values have been cut as well. How many of the most extreme values are cut, is determined by the choice of a parameter δ in the ACER implementation. Here $\delta = 1$, and a lower value will cut more data, and a higher value will include more. This parameter helps the optimization go smoothly, but it is worth noting because it can have an impact on the fitted tail. We also keep in mind that the POT implementation we have used with fExtremes [15], doesn't cut any of the most extreme values.

7.2 Analysis of Returns

After having gone through Byström's approach with a newer data set, and made some brief comparisons between POT and ACER methods, we decided to switch the analysis from using the net returns series to the returns series. As we recall from Equation 3.2 net returns are calculated by

$$N_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{7.3}$$

We observed in Figure 3.1 that the Elspot prices came close to zero in several periods throughout that time series. Our primary concern is that by dividing by P_{t-1} to calculate net returns, we might 'produce' extreme values simply by dividing by very low prices.

	Normal	Student's t
AR-GARCH parameters		
a_0	-0.00852 (0.001)	-0.00852 (0.0012)
a_1	0.148 (0.00355)	0.125 (0.00318)
a_2	0.311 (0.00356)	0.289 (0.00377)
a_3	0.184 (0.00304)	0.216 (0.00342)
ϕ_0	0.0579 (0.00254)	0.21 (0.00615)
ϕ_1	0.236 (0.00476)	0.656 (0.0124)
ϕ_2	0.763 (0.00526)	0.343 (0.00776)
v		3.09 (0.0277)
Standardized residuals statistics		
Mean (%)	0.0285	0.0277
Standard deviation (%)	1.06	1.21
Skewness	1.40	4.75
Excess kurtosis	34.3	232.6
Q(10)	1265	1238
Q(20)	1324	1292
$Q^2(10)$	69.13	12.58
$Q^2(20)$	87.62	17.62
GPD parameters with POT		
ξ	0.297 (0.0234)	
α	0.760 (0.0221)	
u	0.05	

Table 7.4: *AR-GARCH parameters, statistics on the standardized residuals, as well as GPD parameters for the returns series*

The AR-GARCH fit for the returns is shown in Figure 7.4. The AR-GARCH parameters are similar to the fit for the net returns series. In particular we note that the GDP parameters have changed from $\xi = 0.312$ to $\xi = 0.297$ and from $\alpha = 0.884$ to $\alpha = 0.760$.

As we explained in the previous section, we need to pay attention to how we cut the tail in the ACER modeling.

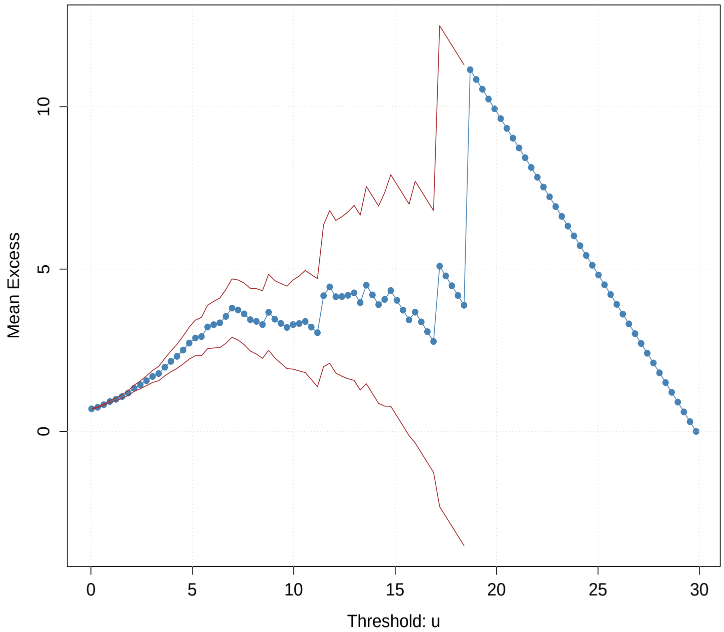

Figure 7.3: *Mean residual life plot for the standardized residuals of the AR-GARCH fit of net returns.*

In Figure 7.3 we have made a *mean residual life plot* of the standardized residuals. A MRL plot can be used as an aid in threshold selection, but gives no definite answers. In the plot we want to look for linearity or intervals of 'regular' behavior. There seems to be a trend up until a value of about seven for the threshold, and a less consistent trend up until about a value of 18 for the threshold.

The ACER functions can also be used to diagnose tail behavior, and in Figure 7.4 we have plotted ACER functions for $k = 1, k = 2, k = 25$ and $k = 169$

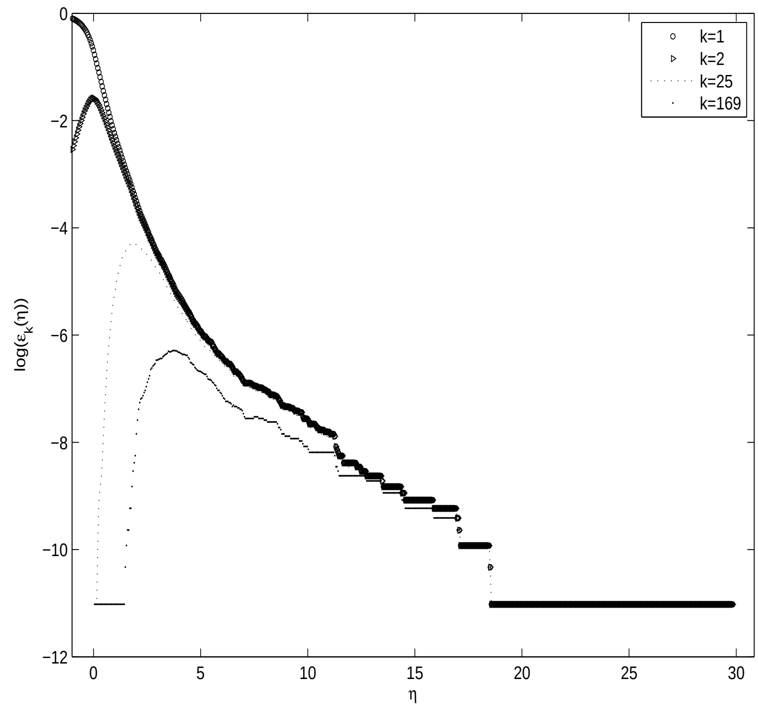

Figure 7.4: *A selection of ACER functions for Elspot price changes, January 1, 2005 to December 31, 2011.*

The plot of ACER functions does not look as 'nice' as the plot of ACER functions in Figure 7.1. By this we mean that the tail bahavior looked to be more regular in the case where we used net returns as input to the AR-GARCH model. We also saw in the QQ-plot in Figure 3.13 of the net returns, that the net returns seemed to be 'stretched' out more nicely in the tail.

Regarding the selection of a threshold we see in Figure 7.4 some of the same things we saw in the MRL plot. But the plot of ACER functions gives us more information since part of the time dependence structure is uncovered. Judging from this plot choosing a good tail marker will be a difficult task because the interval with most regular behavior for ACER_1 (up until a threshold value of seven), is not valid for ACER_{169}.

The drop in the ACER function for $k = 169$ showed that there is still time dependence in the standardized residuals that the AR-GARCH filter didn't pick up.

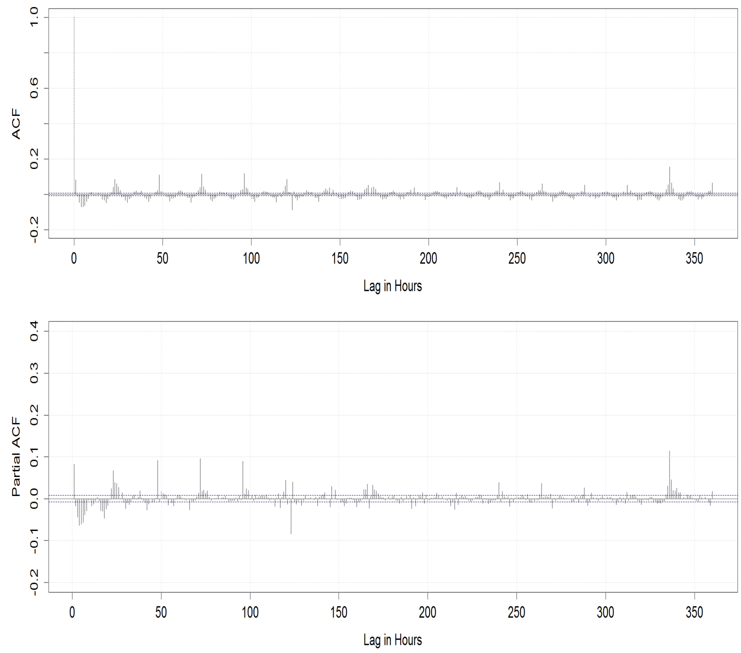

Figure 7.5: *ACF and PACF for the standardized residuals of the AR-GARCH model of returns.*

In Figure 7.5 we have plotted the ACF and PACF of the standardized results to do some extra investigation of the time dependence. Compared to the ACF and PACF in Figure 3.5 of the returns, we can see that alot of the correlation has been caught by the model. In particular the correlation at the weekly lag is now much less pronounced.

We found this a bit odd after observing a significant time dependence at the weekly lag in the ACER functions. But by studying the AR-GARCH model in Equations 7.1 and 7.2 model we found some issues that may explain this observation.

The AR filter is fitted with the weekly lag, but the fit is made with all of the data, so the filter has no chance to predict extreme values. The GARCH model attempts to scale down the extreme values when they come in clusters, but the GARCH model is not fitted with the weekly lag, so it will not scale down extreme values that are further apart.

In the POT method that Byström used in his analysis, the standardized residuals were assumed to be independent since no declustering was done after filtering with the 5.5% threshold. We have uncovered with the ACER method that this may have been a mistake.

ACER function k	q	a	b	c	e	50-year return level
$\eta = 1.59(0.95),\ \delta = 1$						
1	0.143	0.811	0.209	2.53	0.953	71.4 [39.4, 118]
25	-	-	-	-	-	-
169	-	-	-	-	-	-
$\eta = 4.06(0.995),\ \delta = 0.99$						
1	0.00478	0.65	4.11	0.868	0.205	51.4 [25.5, 446]
25	0.00478	0.65	4.11	0.907	0.205	46.6 [4.3, ∞]
169	-	-	-	-	-	-
$\eta = 7.25(0.999),\ \delta = 1$						
1	0.00102	0.0194	6.06	3.06	1.48	67.3 [18.9, 2320]
25	0.000931	0.0224	6.16	3.03	1.60	78.5 [19.2, 4140]
169	0.000548	0.157	7.3	1.5	0.80	79.4 [- , -]

Table 7.5: *Return level estimates with ACER*

By our line of reasoning we would be fast to select the ACER_{169} function, but if we study Figure 7.4 again there are a few problems. Like we have already discussed, there is a difficulty with choosing a good threshold, and this looks to be even harder for the ACER_{169} function. Also we will have to rely on far less data if we use the ACER_{169} function.

Because we saw no clear way to select ACER functions and cut the tail, we decided to make parameter estimates with different setups of ACER functions and tail-cutting parameters, η and δ.

In Table 7.5 we have plotted the parameter estimates together with a 50-year return level estimate. In applications 50 years is not really relevant, but we wanted to see how the different parameter estimates would affect asymptotic behavior. Another thing to note is that the parameter c is equal to one in the asymptotic case, so we should be weary of values that deviate significantly from one. How much significantly is, we are not equipped to say. But values of the c parameter of 3.06 like we observed in Table 7.5 has to raise some concern about overfitting the sub-asymptotic case at the cost of maybe skewing the fit of the shape parameter ξ.

The results in Table 7.5 are hard to make any sense of because there doesn't seem to be much consistency.

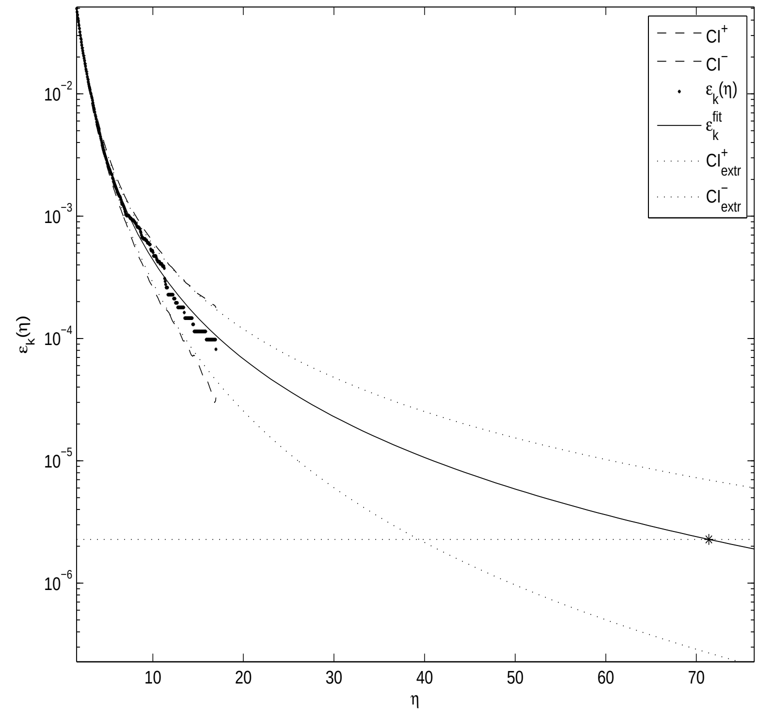

Figure 7.6: *50-yera return level plot for $ACER_1$ with $\eta_1 = 1.59$ and $\delta = 1$.*

In Figure 7.6 we have shown the return level plot of the parameter setup the gave us the tightest confidence bounds. This setup uses the same tail marker at the POT method, but cuts a few of the most extreme observations. So we might expect a similar shape parameter ξ.

We see in Table 7.5 that the shape parameter is 0.953 compared to 0.297 for the POT method. This is a wide discrepancy, but we also see that the c parameter is 2.53, and that we may have over-fitted the sub-asymptotic behavior in this instance. We estimated the same parameter setup (with $k = 1, \eta_1 = 1.56, \delta = 1$) for the net returns to compare. We then found estimates more in line with the POT estimates with $\xi = 0.218$. Here c was 0.891, so that is another indication that we should be skeptical of estimates where the c parameter deviates significantly from one.

ACER function k	q	a	b	c	e	50-year return level
$\eta = 1.59(0.95),\ \delta = 2$						
1	2899	10.2	0.164	0.189	9.5e-5	44.3 [29.6, 63.9]
25	-	-	-	-	-	-
169	-	-	-	-	-	-
$\eta = 4.06(0.995),\ \delta = 2$						
1	0.00465	0.595	4.11	0.873	0.14	40.1 [26.3, 131]
25	0.00355	0.491	4.11	0.905	0.133	39.3 [25.7, 138]
169	-	-	-	-	-	-
$\eta = 7.25(0.999),\ \delta = 2$						
1	0.000992	0.0409	6.49	2.52	1.02	47.8 [- , -]
25	0.00921	0.0423	6.41	2.43	0.986	48 [22.5, 345]
169	0.000547	0.0286	6.23	2.42	1.02	49.8 [- , -]

Table 7.6: *Return level estimates with ACER*

In Table 7.6 we have tried experimenting with a δ parameter equal to 2, to see what happens when all of the most extreme observations are included.

7.3 Forecasting Extreme Values

Our goal in this thesis is not really to predict when extreme values will occur, but to say something about the probability that a given threshold will be exceeded within a certain time frame.

Predicting quantiles is done by using Equation 6.3

$$\alpha_{t,p} = a_0 + a_1 r_{t-1} + a_2 r_{t-24} + a_3 r_{t-168} + \sigma_t \alpha_p \tag{7.4}$$

where α_p are quantiles from the extreme value distributions we have fitted. Our predicted quantiles from the fitted distributions will be scaled by the GARCH model and then added to the trend predicted by the AR filter.

This will give us predictions for the returns series, but we want to see what happens with the predictions for the Elspot prices.

We recall from Equation 3.1 that returns were found as

$$R_t = P_t - P_{t-1}. \tag{7.5}$$

This also mean that the next expected price at time t would be the current price plus the expected return

$$E[P_{t+1}] = P_t + [R_{t+1}] \tag{7.6}$$

We were curious what would happen in periods with large price differences.

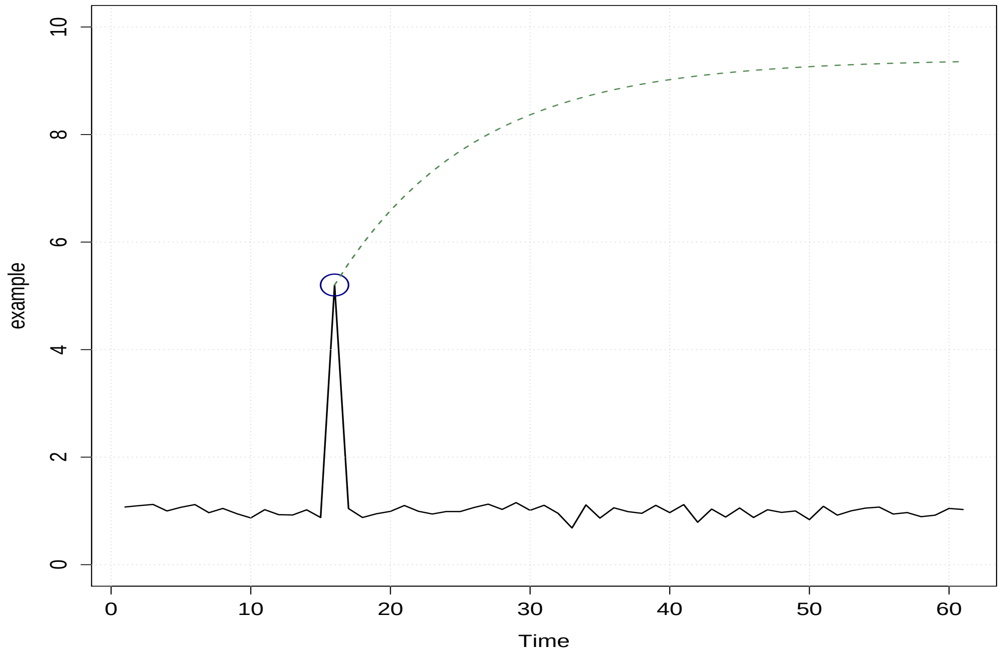

Figure 7.7: *Constructed example of forecasting after a spike (forecast dotted in green).*

In Figure 7.7 we have tried to illustrate what might happen with predictions after a spike. The return at time t will be extremely large. This value will then be used in Equation 7.7 to calculate the expected return at time $t + 1$

$$E[r_{t+1}] = a_0 + a_1 r_t + a_2 r_{t-23} + a_3 r_{t-167}. \tag{7.7}$$

Without doing the exact calculations we can then imagine that the result will be something like the plot in Figure 7.7 if we continue to calculate expected returns without using information about the actual price past time t.

We wanted to study this further.

Chapter 8

Returns & Extreme Values

When we model returns we have to make some assumptions about their dependence and distribution. If we assume iid normal returns we are working under the assumptions of the *Wiener process*

$$W_t = W_t - W_s \sim \mathcal{N}(0, t - s) \tag{8.1}$$

where W_t is a time series and $t - s$ is the lag difference. In the case of a returns series, s would be the first lag, $t - 1$, at time t.

A modified version of the Wiener process, where the properties of the process have been changed so that there is a tendency of the process to move back towards a central location, with a greater attraction when the process is further away from the centre, is called a *Ornstein-Uhlenbeck process*

$$dx_t = \theta(\mu - x_t)dt + \sigma dW_t \tag{8.2}$$

with solution given by

$$x_t = x_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \int_0^t \sigma e^{\sigma(s-t)} dWs \tag{8.3}$$

The Ornstein-Uhlenbeck process can be thought of as the continuous-time analogue of the discrete-time AR(1) process.

Both of these processes have normally distributed noise as their input. We tried to make some assessments about what happens when we take returns of Elspot prices, and in particular the extreme values.

8.1 A Stock Market Comparison

For a series of stock prices it makes intuitively sense to take returns because the stock price always reflect the underlying asset, the company's value. And if the stock market is *efficient*, then at no point in time should the stock's *price history* influence the future prices of the stock. I.e. there should be no *arbitrage* opportunities in the sense that you should not be able to tell anything about the future price development of a stock based on recent price jumps or price falls. Therefore, if you have a time series of stock returns, the returns should be independent and identically distributed

These assumptions may be harder to make for the Elspot returns. In Section 2.2 we explained how the Elspot market works. We recall that bid and ask offers are placed on one-hour-long contracts, and that the bids are placed each day at noon for the next 24 hours, starting at midnight. In such a market there is no asset or entity that changes its value from one hour to the next. Electricity is bought by the hour and consumed by the hour.

In Section 3.3 we calculated ACFs for 02:00-03:00 and 08:00-09:00 hours, and visualized their differences in Figure 3.8. Clearly the hours' prices behave differently, so the perceived entity that changes value from one hour to the next, is not really the same thing.

8.2 Mean Reversion

When we chose to model the returns with an AR-process we assumed a priori that the returns were mean reverting without thinking about it. We performed an R/S analysis on the returns to get a better idea of the mean reversion in the hourly Elspot prices.

Rescaled Range Analysis

This method, described in Rafael Weron's book *Modeling and Forecasting Electricity Loads and Prices* [13], allows for the calculation of the self-similarity parameter H, which measures the intensity of long-range dependence in a time series [13].

The analysis begin with dividing a time series (of returns) of length L into d subsets of length n. Next for each subseries $m = 1, ..., d$:

1. find the mean (E_m) and standard deviation (S_m);
2. normalize the data $Z_{i,m}$ by subtracting the sample mean $X_{i,m} = Z_{i,m} - E_m$ for $i = 1, ..., n$;
3. create a cumulative time series $Y_{i,m} = \sum_{j=1}^{i} X_{j,m}$ for $i = 1, ..., n$;
4. find the range $R_m = \max\{Y_{1,m}, ..., Y_{n,m}\} - \min\{Y_{1,m}, ..., Y_{n,m}\}$;
5. rescale the range R_m/S_m

Finally, calculate the mean value $(R/S)_n$ of the rescaled range for all subseries of length n.

It can be shown that the R/S statistic asymptotically follows the relation $(R/S)_n \sim cn^H$. Thus the value of H can be obtained by running a simple linear regression over a sample of increasing time horizons

$$\log(R/S)_n = \log c + H \log n. \tag{8.4}$$

Equivalently, we can plot the $(R/S)_n$ statistic against n on a double-logarithmic paper. If the returns process is white noise then the plot is roughly a straight line with slope 0.5. If the process is *persistent* then the slope is greater than 0.5; if it is *anti-persistent* (or *mean reverting*) then the slope is less than 0.5.

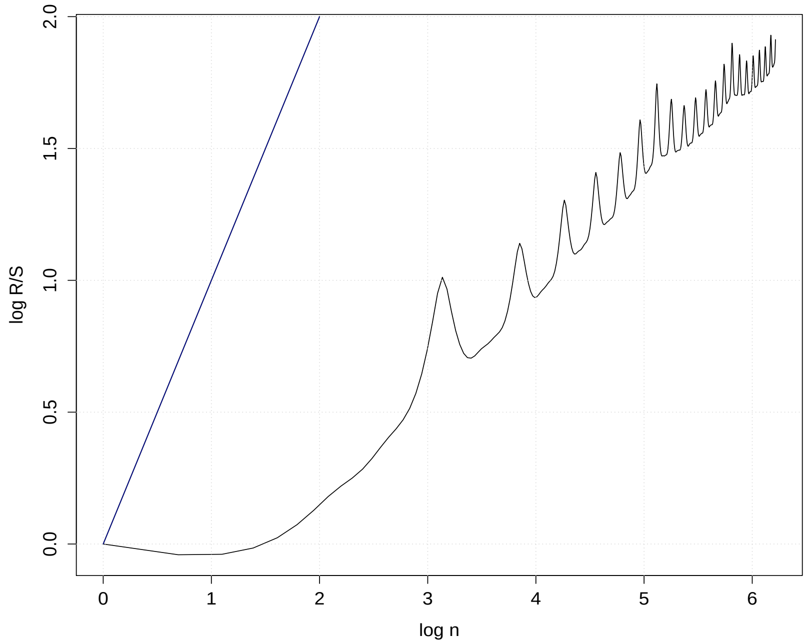

Figure 8.1: *R/S analysis plot. The slope of the curve estimates the Hurst parameter. The blue line is for a Hurst parameter of 0.5.*

In Figure 8.1 we can clearly see that the Hurst exponent, or self-similarity parameter is well below 0.5, which indicates that the returns series is mean reverting.

But other than confirming that our time series is mean reverting, the Hurst exponent does not say us all that much. We tried to make some plots to help us understand more about how the time series was mean reverting. We defined the sums

$$\Delta_t = \sum_{i=0}^{l-1} R_{t-i} \tag{8.5}$$

$$\nabla_t = \sum_{i=1}^{l} R_{t+i} \tag{8.6}$$

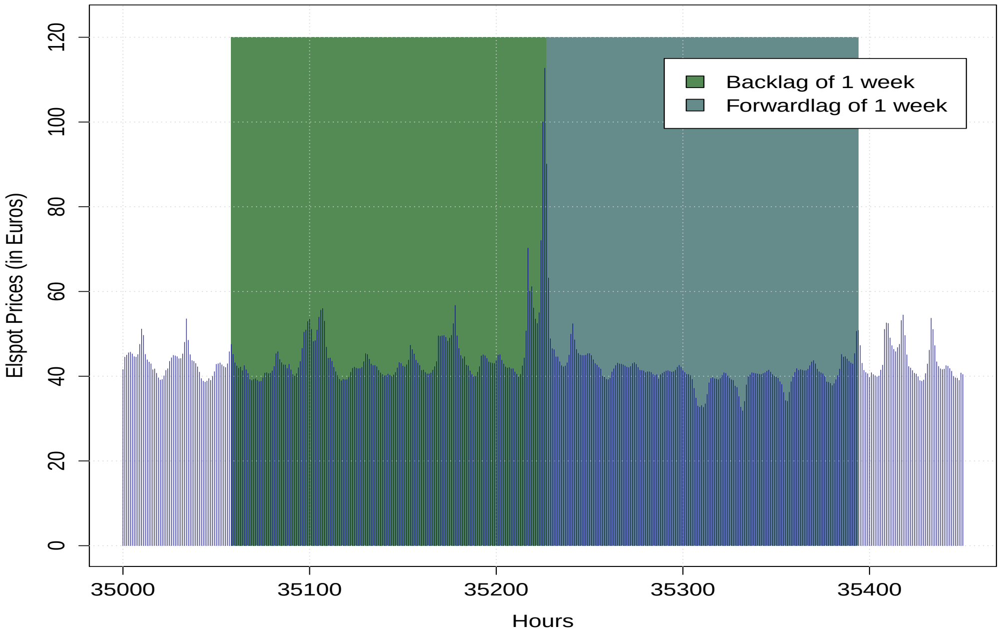

Figure 8.2: *Excerpt from the Elspot prices series showing how*

Figure 8.2 visualizes how these sums are calculated at time t.

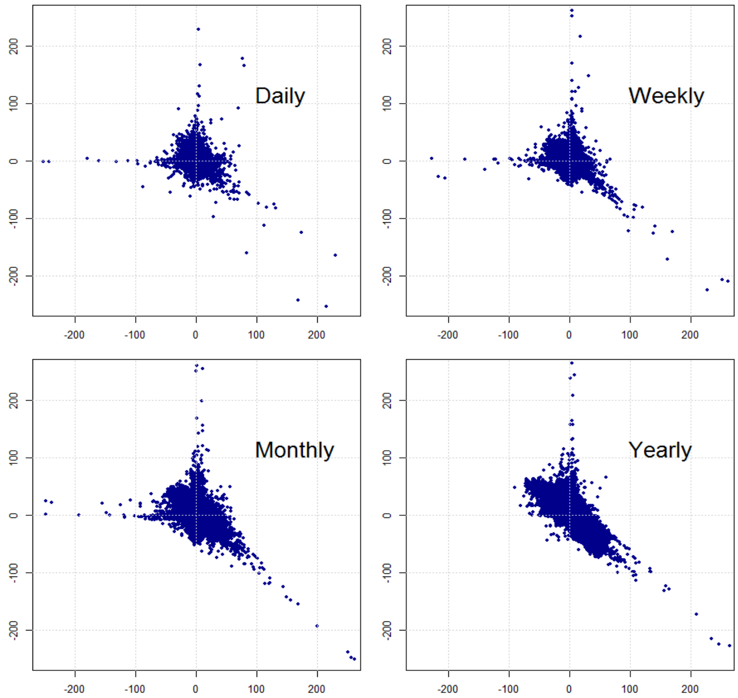

Figure 8.3: ∇_t *against* Δ_t *with lags of a day, a week, a month and a year.*

In Figure 8.3 we have plotted ∇_t against Δ_t in a scatterplot, with $l = 24, 168, 720, 1248$, with t running through the entire time series. The plots show that if we have a period where the sum of returns is very high, i.e. the price has risen, the next period of equal length will tend to have an equally large sum with opposite sign.

This again indicates mean reversion, but we were wondering what the time series was mean reverting to, and how fast it was happening.

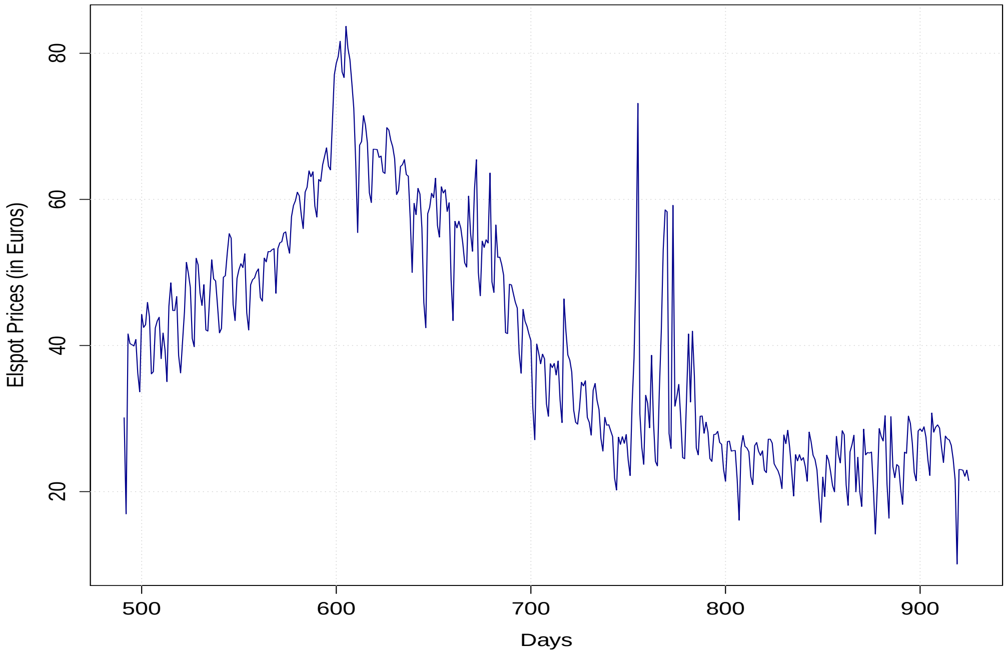

Figure 8.4: *Excerpt from daily Elspot prices at 08:00-9:00*

In Figure 8.4 we have shown an example from the hourly Elspot prices where the trend has as much variation as the spikes. In such a scenario it is difficult to seperate a spike from a trend. In Figure 8.2 the trend is much easier to identify, although we do note that even in that figure, the spike is not a singular value, but the sum of several big increases.

8.3 Asymmetric Transformation of Extreme Quantiles

In Section 7.3 we ran into the problem of forecasting after observing a spike. We constructed another example here to highlight an issue with taking returns.

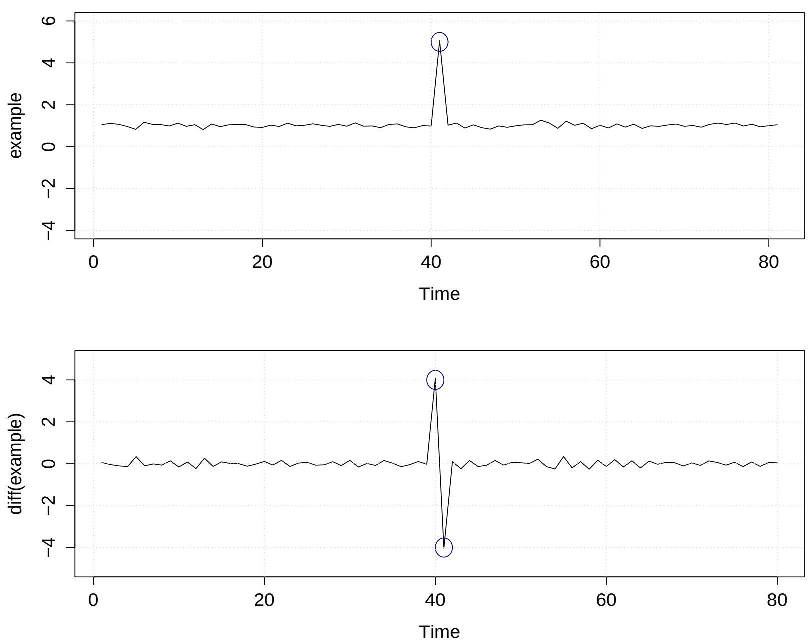

Figure 8.5: *Constructed example of a time series with a single spike, and its differenced series*

If we happen to have a large singular spike like in Figure 8.5. The differenced series, or the returns, will now contain two spikes instead of one.

Chapter 9

An Alternative Transformation

Taking returns or differencing a time series is a standard way to make time series more stationary, so we naturally took that approach here aswell. As we discussed in the previous chapter, there can be some issues with the returns series, in particular when analyzing extreme values.

To avoid what we perceived as problems with the returns series we thought about applying ACER directly to the Elspot prices. The problem with this is shown in Figure 9.1.

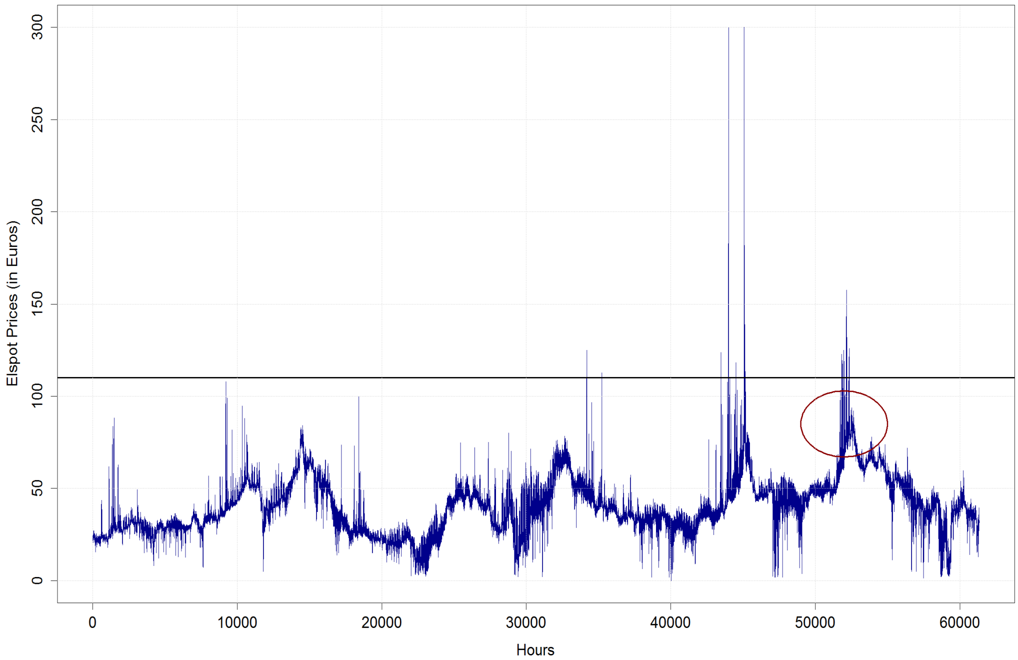

Figure 9.1: *Elspot prices, January 1, 2005 to December 31, 2011.*

To avoid classifying data as spikes that are part of trend, we would have to use a very high threshold. In Figure 9.1 we have circled out where the trend has its highest values, and drawn a line well above this as an approximate threshold. With such a high threshold we can see from the figure that we would have to scrap most of the data and only capture a small percentage of the spikes.

9.1 A Moving Median

Our naive approach was to consider something in-between Elspot prices, and Elspot returns. By forming a time series of differences between the Elspot prices and a *moving median* we hoped to capture all the spikes but at the same time not let the extreme quantiles be distorted by putting too much emphasis on immediate history. We set the moving median to be the median of one week of Elspot prices, or an hourly lag of 168. The differences series becomes

$$D_t = P_t - \text{median}([P_{t-168}, P_{t-1}]) \tag{9.1}$$

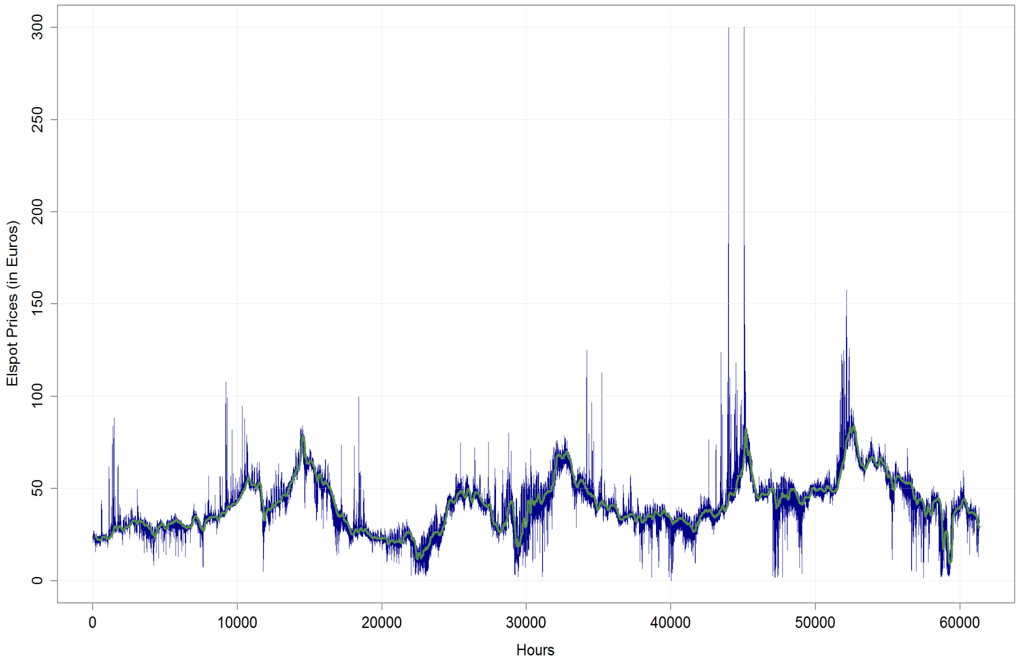

Figure 9.2: *Elspot prices, January 1, 2005 to December 31, 2011 with the moving median in green.*

This method of transforming the data shows some promise after studying Figure 9.2. The moving median seems to follow the trend nicely without being affected by the spikes. There is an obvious issue of what the lag of the moving median should be, but we found that with a lag of one week, the moving median was following the trend fairly close, but at the same time not dipping down, or jumping up with the spikes.

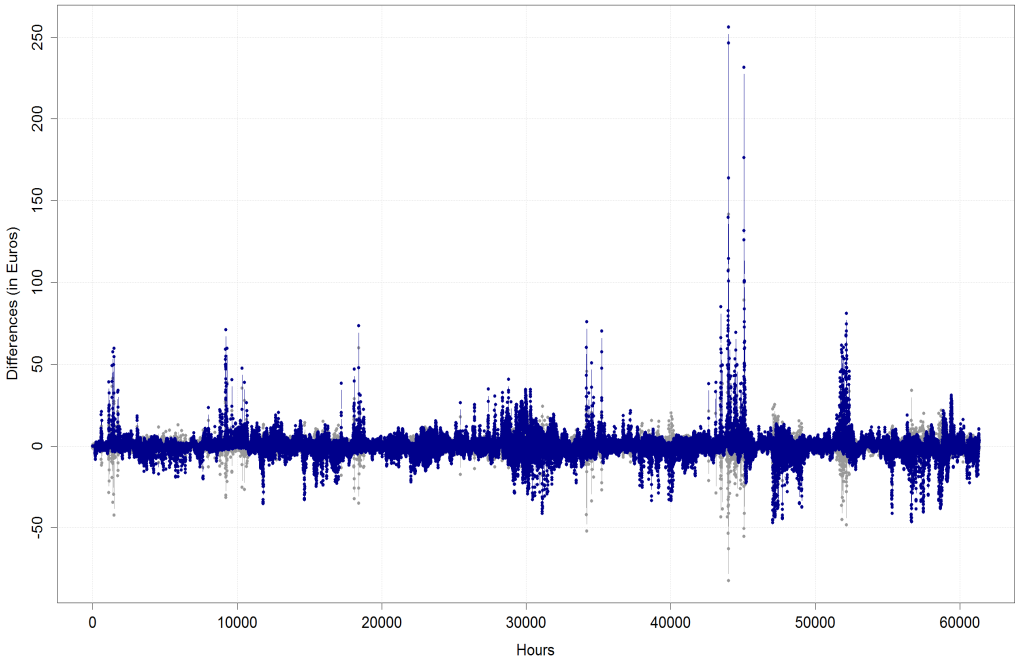

Figure 9.3: *Elspot prices differences with lag according to the moving median. Plotted together with the Elspot returns series in grey.*

By comparing with Figure 3.3 (in grey in Figure 9.3) we see that the series of differences D_t does not have the same symmetry about the x-axis that the return series has. Our constructed example illustrated in Figure 8.5 showed how taking returns could possibly 'produce' new extreme values with some symmetry, so we consider it a good thing that we see less symmetry here.

In the sense that we can consider the returns or differences a residual series themselves, we computed their sum and sum of squares to gauge some of their properties. We see in Table 9.1 that we end up with a much larger squared sum of differences than squared sum of returns.

Data	Sum	Sum of Squares
Returns	5.23	46.8×10^4
Differences	2407	278×10^4
Adjusted differences	10080	247×10^4

Table 9.1: *Squares*

If we wanted to fit a time series model to the differences, then the numbers from Table 9.1 would not be a good sign. Our concern is with the very high quantiles, so this might not matter for us.

9.2 Daily Differences

Kai Erik Dahlen analyzed daily prices in his master's thesis [3], where he looked at the prices for hours 08:00-09:00. Figure 3.7 indicates different behavior for different hours, so we wanted to get a better idea about how different hours differed in distribution.

So we will take the series of differences and divide it up into series for each hour, and look at their empirical distributions.

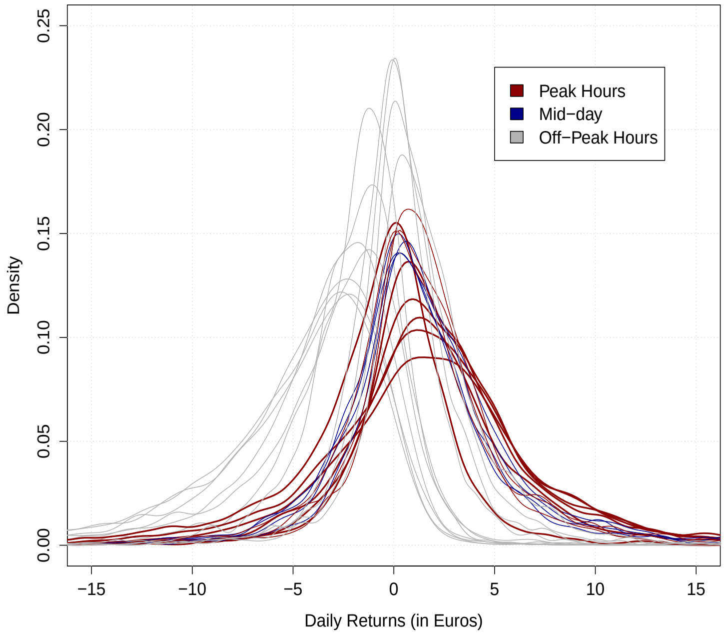

Figure 9.4: *Empirical densities of daily differences for the different hours of the day.*

From Figure 9.4, we can see that the densities do not share a common mean or are centered about a common peak density. This was expected because of the way we constructed the differences series. It might also partly have contributed to the larger numbers for the differences in Table 9.1.

We took the freedom to simply subtract the mode from each distribution to sort of compensate for the daily cycle of prices. We will justify this by admitting that our approach is experimental and more of a practical solution than anything else.

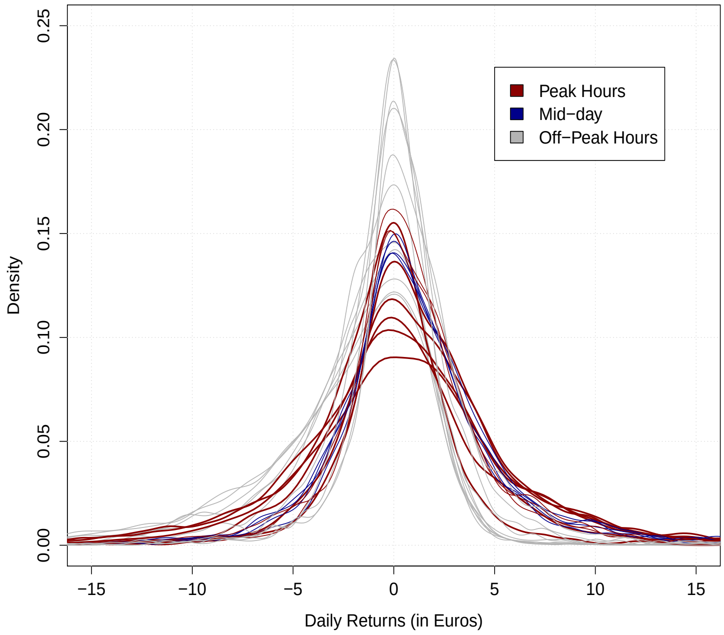

Figure 9.5: *Empirical densities of daily differences for the different hours of the day, each one adjusted to a mode equal to zero.*

After studying Figure 9.5 we realized that the hourly differences contains values from all of these distributions combined, so the distribution of hourly differences can be thought of as the distribution of a sum of different stochastic variables, $X_1, X_2, ..., X_{24}$.

This makes a strong argument for considering daily prices, or in our case, daily differences. The downside is that daily series have far less data. In the next chapter we will take a look at both.

Chapter 10

ACER Analysis of Differences

After making a sort of blind folded transformation of the data in the previous chapter, we applied ACER to the differences to see if this new way of looking at the data has any merit. Because of the different distributional properties of the different hours we looked at both the hourly differences, as we did in Chapter 7 following Byström's approach, and the the daily differences. The daily differences will still be the values from the hourly differences, but because the moving median remains fairly stable, these values will be very similar to separately daily series with the own moving median etc. We tried that, and actually found this method favorable because the moving median was less stable when considering daily data.

10.1 Daily Differences

Even in the data description we saw in Figure 3.7, we saw that the electricity prices of different hours behaved differently. We have looked into them in more detail in this section.

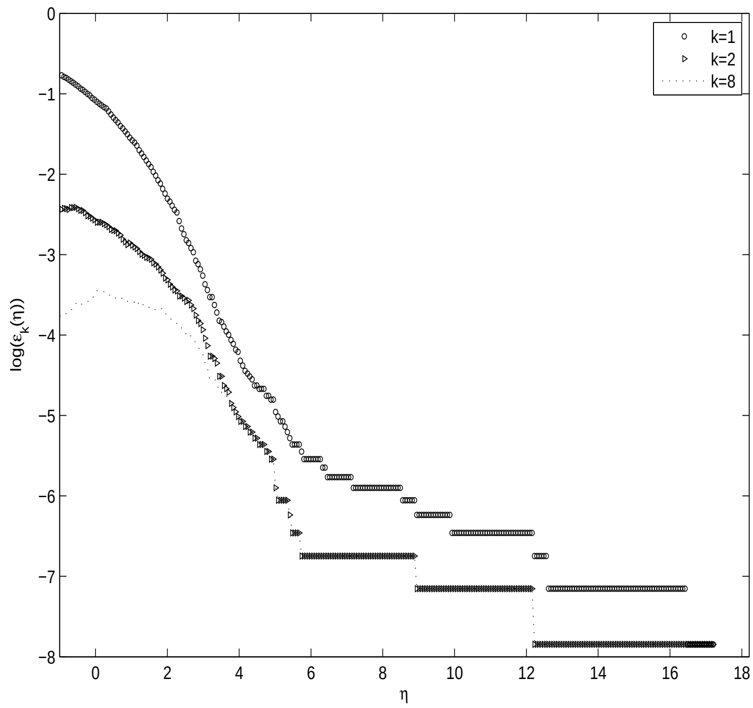

Figure 10.1: *A selection of ACER functions for the differences at 02:00-03:00*

Figure 10.1 shows the ACER functions for independent data, daily, and weekly lag for hours 02:00-03:00. For ACER_1 it looks like we have and exponential tail-off. For ACER_8 the tail off seems more linear, except for the most extreme values.

In Figure 10.2 we have plotted two different fits, one of ACER_1, and of ACER_8, together with their return level plots. The parameter and return level estimates are presented in Table 10.1.

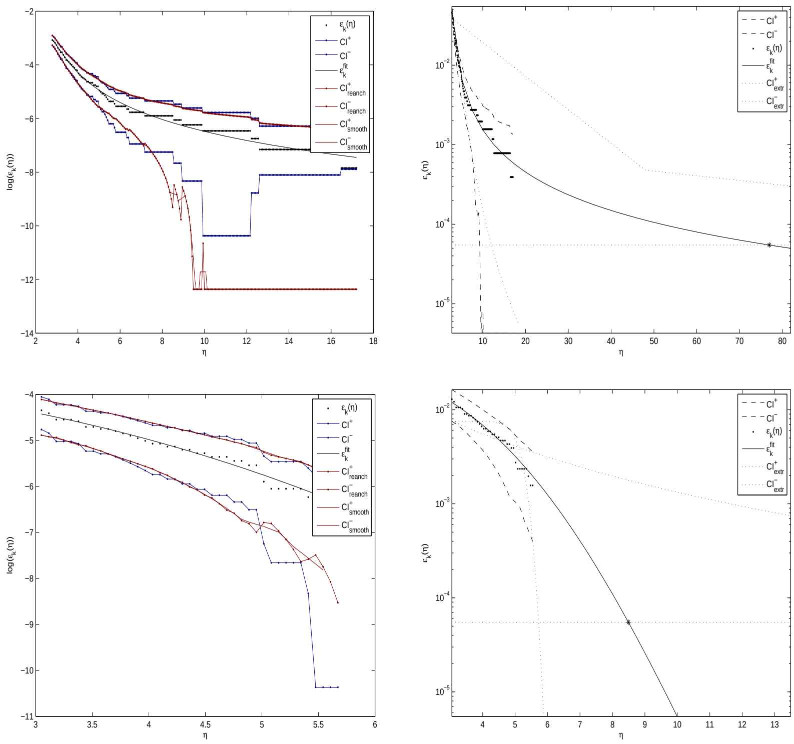

Figure 10.2: *Fitted ACER functions and return level estimates for a selection of ACER function with a selection of tail cuts 02:00-03:00*

The plot shows us that what ACER function we choose and how we cut the tail can have a big impact on the fit. In Table 10.1 we see that the shape parameter ξ tells two very different stories. While the first is clearly a Frechet, the second one looks to be more Gumbel.

ACER function k	q	a	b	c	e	50-year return level
$\eta_1 = 2.73(0.95),\ \delta = 2$						
1	0.0556	1.65	2.35	2.49	1.7	76.9 [12, 515]
$\eta_1 = 3,\ \delta = 1$						
8	0.0376	4.7e-5	-6.54	4.49	0.097	8.49 [5.72, 43.3]

Table 10.1: *Parameter and return level estimates for 02:00-03:00.*

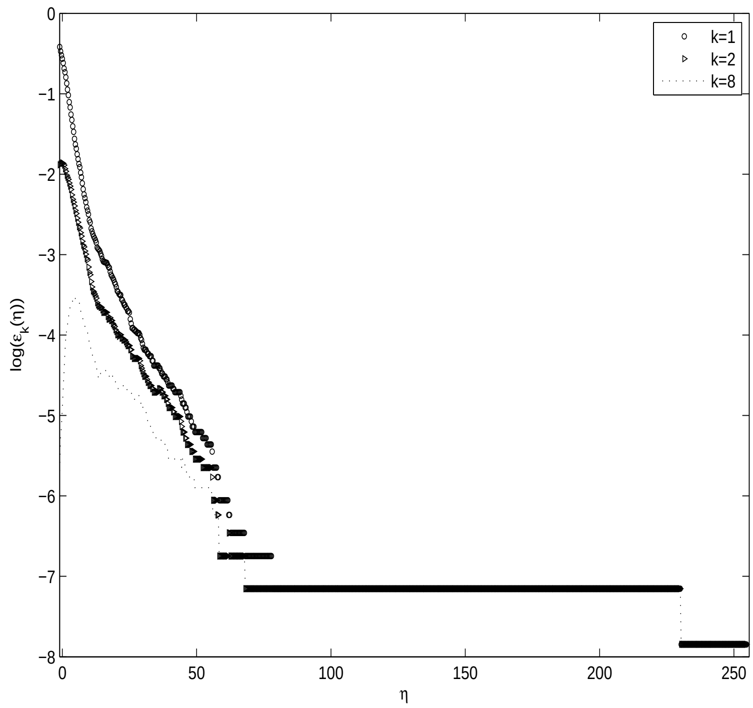

Figure 10.3: *A selection of ACER functions for the differences at 08:00-09:00*

Figure 10.3 shows the ACER functions for independent data, daily, and weekly lag for hours 08:00-09:00.

In Figure 10.4 we have plotted two different fits, one of ACER_1, and of ACER_8, together with their return level plots. The parameter and return level estimates are presented in Table 10.2.

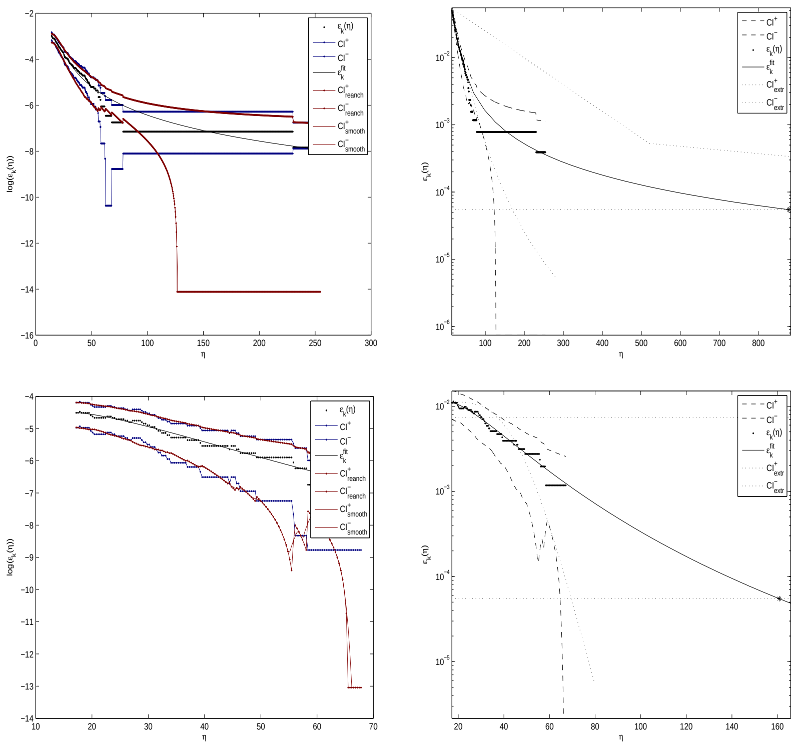

Figure 10.4: *Fitted ACER functions and return level estimates for a selection of ACER function with a selection of tail cuts 02:00-03:00*

In Figure 10.4 and Table 10.2 the same pattern as for hours 02:00-03:00, where the independent with the entire tail, seems to be more heavy-tailed.

ACER function k	q	a	b	c	e	50-year return level
$\eta_1 = 14(0.95),\ \delta = 2$						
1	0.0465	0.0084	14	2.05	1.4	877 [169, 6075]
$\eta_1 = 17,\ \delta = 1.2$						
8	0.011	0.013	17.2	1.4	0.32	160 [69, -]

Table 10.2: *Parameter and return level estimates for 08:00-09:00.*

10.2 Hourly Differences

The hourly differences are more similar to the return series we studied in Chapter 7. As we can see from Figure 10.6 the first ACER function looks to behave nicely with what looks like a smooth curve for the majority of the tail. For ACER_{169} the function is far less regular.

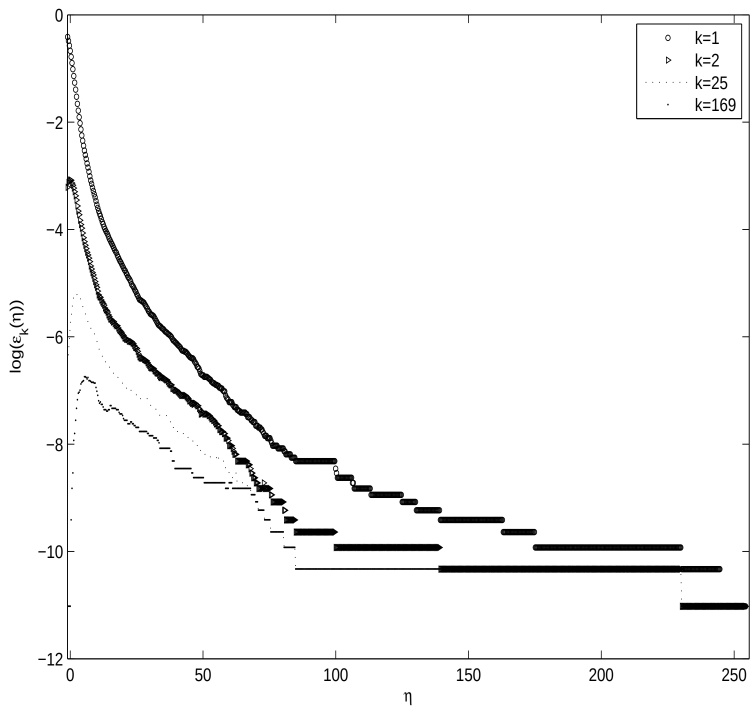

Figure 10.5: *A selection of ACER functions for the differences.*

A thing to keep in my when it comes to the ACER_{169} function is that we dont know *which* exceedances are removed. If we think about the hourly differences as the compilation of daily differences, then the counted exceedances will probably come from different daily series, depending on the size of spikes, their ordering, and threshold. Since these have different distributions, this might be an argument for using the ACER_1 function.

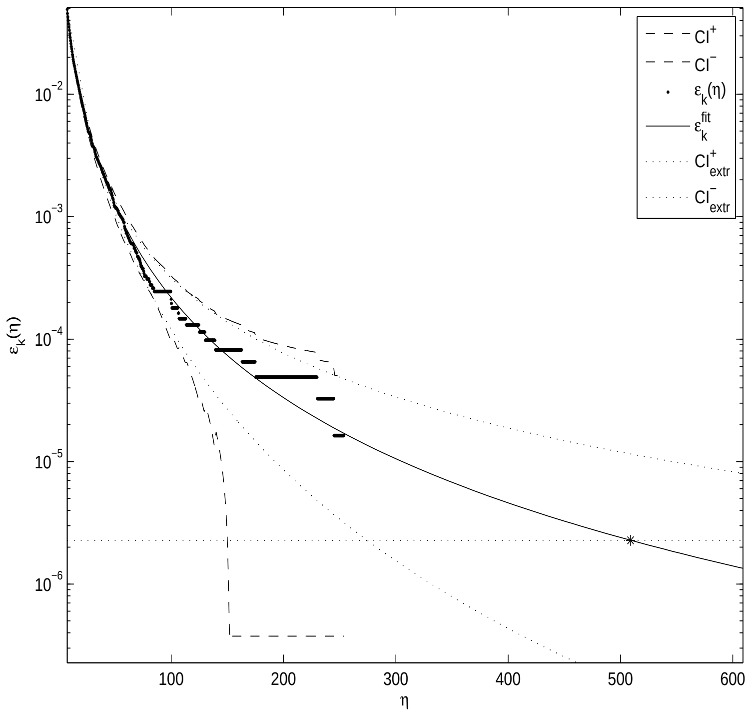

Figure 10.6: *50-year return level plot for ACER*$_1$ *with* $\eta_1 = 7.32$ *and* $\delta = 2$.

In Figure 10.6 we have plotted the return level estimates for ACER_1. It is tempting to use this ACER function because of the nice fit and more reasonable confidence bounds.

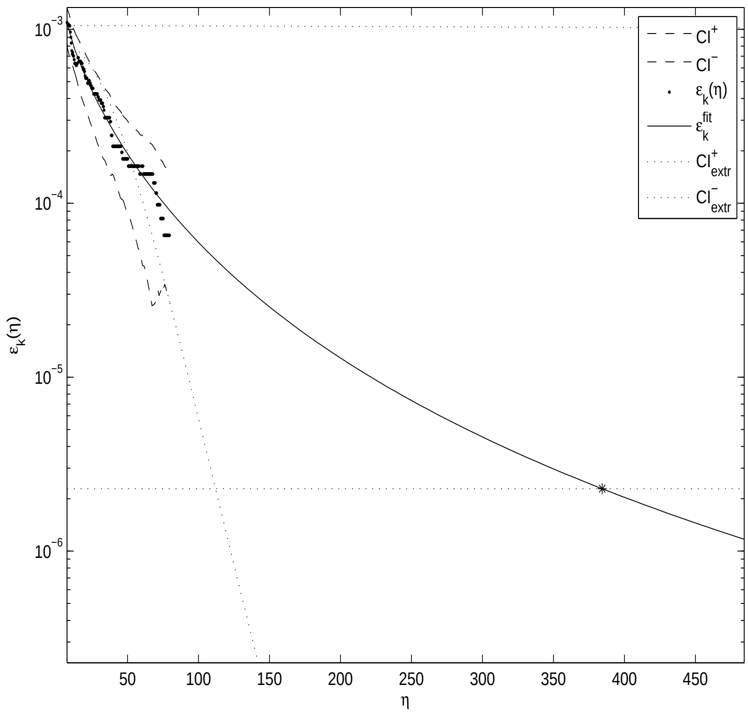

Figure 10.7: *50-year return level plot for $ACER_{169}$ with $\eta_1 = 7.32$ and $\delta = 1$.*

From the ACER_{169} we expected a more gumbel type of fit. But Figure 10.7 shows us some exponential decay in the tail. But as indicated by the confidence bounds, this fit is much less certain.

In Table 10.3 we can see that the parameter and return level estimates are not that different. The shape parameter for ACER_1 is 0.29 and for ACER_{169} it is 0.244. Oddly enough the first ACER function has a shape parameter almost identical to the shape parameter Byström found in his analysis.

We also note that the other parameter estimates look reasonable, i.e. the c parameter is close to one, which we consider a good thing.

ACER function k	q	a	b	c	e	50-year return level
$\eta_1 = 7.32(0.95),\ \delta = 2$						
1	0.0493	0.231	7.37	0.89	0.29	509 [279, 1147]
$\eta_1 = 7.32,\ \delta = 1$						
169	0.00107	0.0812	7.37	0.87	0.244	384 [112, 183185]

Table 10.3: *Parameter and return level estimates for hourly differences.*

Table 10.3

10.3 Forecasting With Differences

To make prediction with the approach we have used here, we can select a probability from a distribution we have fitted, and then add the corresponding quantile to the moving median at time t.

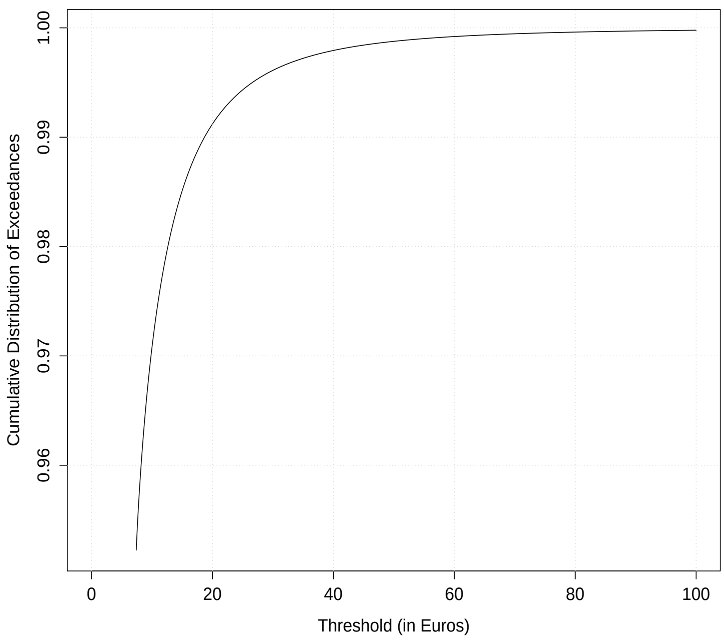

Figure 10.8: *Cumulative distribution function of exceedances with the fit of the $ACER_1$ function of hourly differences. The tail starts at 7.37 (0.95).*

In Figure 10.8 we have shown an example of a fitted distribution. To illustrate how we can use this distribution we applied it to calculating the probability of the electricity price exceeding 70 Euros, at the price where the Elspot time series we have studied ends.

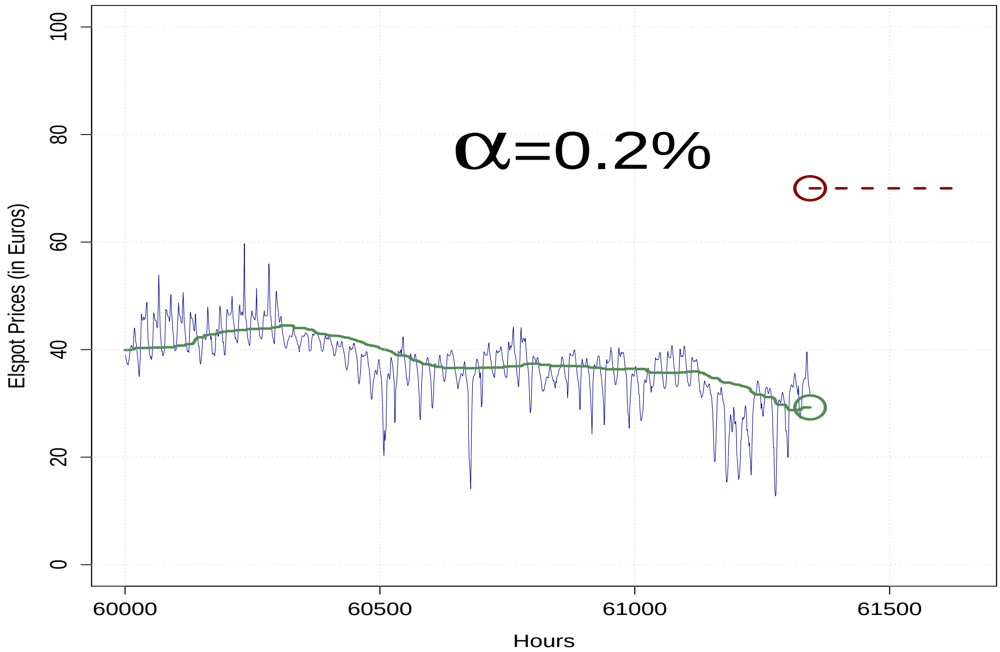

Figure 10.9: *An excerpt from the end of the Elspot prices series with the moving median in green. α is the probability of exceeding the threshold dotted in red.*

At the very end of the time series the price is 31.46 Euros, and the moving median is 29.25 Euros. To find the probability of exceeding 70 Euros we would then need to find the probability from our fitted distribution corresponding to a quantile of $70 - 29.25 = 40.75$. The result is shown in Figure 10.9. In comparision the probability of exceeding 70 Euros, if we fitted a normal distribution to the hourly differences, would be ≈ 0.

We could relatively easily improve these estimates by including some scaling depending on what hour the price is for, and to also include some dependence on recent volatility.

Chapter 11

Discussion & Concluding Remarks

We started the work with this thesis without much prior knowledge about electricity prices or how to model them. But after learning about Nord Pool, how the electricity prices are set, and about the many ways that electricity prices can be modeled, our understanding has matured greatly.

Our task was to study extreme prices and their statistical properties, but much of our time was consumed by studying and thinking about different ways to model and predict electricity prices. There has been written a lot about this subject and electricity prices have been, and are being modeled in many different ways. It seemed like there was no 'right' way to approach the modeling problem.

A realization came to us after several talks with Eirik Mo at Statkraft. Because he explained that the fundamental models that Statkraft uses can actually predict even extreme prices fairly well. At first this was disheartening because it seemed like it took away the entire motivation to model extreme prices. But the simple, yet important detail to remember, was the *time frame* of the predictions.

The time frame of a prediction also determines how much relevant information is available for that prediction. So the more information we have about, for instance, consumption, weather, reservoir levels, etc, the better predictions we can make for the next hour, day and so on. But if we go much further in time, let us s say a year, then it would not be very useful to know these things anymore. So put in another way, there is more *uncertainty* the less we know about influencing factors, and the further in time we are trying to predict.

Trivial as it really is, this had an important implication for us. The stochastic properties of the noise may be different depending on what information is being used and what time frame is considered. In many applications and regression methods, noise is often assumed to be Gaussian. For electricity prices we were assuming more fat-tailed behavior from the get-go, but did not really think about *what* had this fat-tailed behavior. Because we can not really talk about fat-tailed behavior in electricity prices without defining a stochastic variable of some sort. And as we have seen in this thesis, the stochastic variable is usually some residual from a model that has both transformed and trend-fitted the prices.

In the case with Statkraft, who has models that can predict extreme prices fairly well, the models residuals does probably *not* have fat-tailed behavior. So we found ourselves wondering why we were even applying extreme value theory in the first place. The answer to that has really been the ongoing process of working with this thesis.

Our opinion is that it's a matter of how the data is approached, transformed, and modeled. And also what the models and predictions are trying to achieve. The SUR regression model is a good example of a model that wanted to take full advantage of the correlation structure both between days and hours to predict the trend as well as possible in the short term (10 days). Statkraft's models are even more advanced and make even better predictions. We found that our goal was not to create a model that could make good predictions, but rather find a way to say something about the distribution of extreme prices. We did this by using an AR-GARCH time series model inspired by Byström, and found similar results as him with a sample of newer data.

Additionally, we believe we found some shortcomings in the way Byström transformed his data. Although his method manages to describe the residual series' very high quantiles fairly well, it was not clear to us how the distributions he fitted related to the prices themselves, as that was made no mention of. We tried to approach the data in a different manner, to hopefully be able to relate extreme value distributions to the electricity prices in a more intuitive and applicable way, and to some degree we believe that this has been achieved.

In our extreme value analysis the ACER approach has been a great asset in getting a more detailed picture of what happens in the tail of the fitted distributions, and also how the introduction of extra parameters can be used to fit sub-asymptotic behavior more appropriately when there is not enough data to make a good fit of an asymptotic distribution. To analyze time dependences in extreme prices ACER has also proved very helpful.

Some other problems that are of a more inherent type, is the sample size and the general applicability of statistics to analyze extreme electricity prices. In our data analysis we found many cases where there were simply not enough data in the tail to make confident estimates, particularly when we looked at daily data. As we understand this is a recurring problem in extreme value statistics, and is just 'the nature of the beast' so to speak.

Regarding extreme value statistics' applicability to describe extreme electricity prices, we found several instances where it seemed hard to identify any regular pattern in the tail behavior. We must remember that it is not a given that with whatever we consider a stochastic variable, there will be a distribution that describes it well. If we think about, we would almost expect the opposite is true since the Elspot market is continually changing and evolving. This also goes for the market players who must try to adapt their bidding to the best of their abilities so that they can avoid over-paying for electricity, or in the case of a supplier, not getting as high a price as they could have for their power.

Bibliography

[1] Hans N.E. Byström. Extreme value theory and extremely large electricity price changes. *International Review of Economics and Finance*, 14:41, 2005.

[2] Stuart Coles. *An introduction to Statistical Modeling of Extreme Values.* Springer, 2001.

[3] Kai Erik Dahlen. Comparison of ACER and POT methods for estimation of extreme values, Master's thesis NTNU, 2010.

[4] Alexios Ghalanos. *rugarch: Univariate GARCH models.*, 2012. R package version 1.0-9.

[5] Richard Harris and Robert Sollis. *Applied Time Series Modelling and Forecasting.* John Wiley and Sons, 2003.

[6] MATLAB. *version 7.12.0.635 (R2011a).* The MathWorks Inc., Natick, Massachusetts, 2011.

[7] Alexander J. McNeil and Rüdiger Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7:271, 2000.

[8] Arvid Næss. Estimation of extreme values of time series with heavy tails. *Preprint Statistics, NTNU*, 14, 2010.

[9] Nord Pool. Price calculation. http://www.nordpoolspot.com/How-does-it-work/Day-ahead-market-Elspot-/Price-calculation/.

[10] Nord Pool. History. http://www.nordpoolspot.com/About-us/History/, 2012. [Online; accessed 10-Jun-2012].

[11] R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[12] Kim Stenshorne. A framework for constructing and evaluating probabilistic forecasts of electricity prices, Master's thesis NTNU, 2011.

[13] Rafal Weron. *Modeling and Forecasting of Electricity Loads and Prices.* John Wiley and Sons, 2006.

[14] The Facts Wind Energy. The nordic power market - nordpool spot market. http://www.wind-energy-the-facts.org/fr/part-3-economics-of-wind-power/chapter-5-wind-power-at-the-spot-market/power-markets.htm, 2012. [Online; accessed 10-Jun-2012].

[15] Diethelm Wuertz, many others, and see the SOURCE file. *fExtremes: Rmetrics - Extreme Financial Market Data*, 2009. R package version 2100.77.

[16] Øyvind Lie. Ikke pristak på strøm. http://www.tu.no/energi/2010/12/15/ikke-pristak-pa-strom, 2010. [Online; accessed 14-April-2012].