Mathematics MSc Dissertation MTHM038, 2019/20

On-off Intermittency in Spot Price Market Data

Elliz Akindji, ID 160412097

Supervisor: Prof. Wolfram Just

A thesis presented for the degree of
Master of Science in *Mathematics*

School of Mathematical Sciences
Queen Mary University of London

Declaration of original work

This declaration is made on September 11, 2020.

Student's Declaration: I, Elliz Akindji, 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 italic fonts, **and**
2. using quotation marks "...", **and**
3. explicitly mentioning the source in the text.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Wolfram Just, without whom I would not have been able to complete this project. I am extremely grateful for his assistance during every step of this research project, his willingness to offer his time so generously has been truly appreciated along with his tremendous patience and constant support.

Abstract

In this paper, we study on-off intermittency behaviour in the case of electricity spot market prices observed in the Nord Pool. For this purpose, we analysed the returns of spot market prices. High and low peaks reveal the existence of extreme events – our main ambition in this thesis is to investigate in what way extreme events influence the cost of electricity. The task is to determine if the spot prices comply with on-off intermittency behaviour. On-off intermittency in dynamical systems requires the exhibition of two distinctive states during the course of the time series; the "on" and "off" states. The "off" state being where the quantity stays nearly constant compared to the "on" state, where the quantity has a fleeting burst away from the constant value. Work has been previously presented to reveal a universal scaling behaviour for such on-off intermittent systems; the length of time intervals between bursts should follow a $-3/2$ power law distribution. Within our spot price data, we discover price spikes and volatility, investigate the differences in hours of extreme prices and attempt to fit a power law. The standard way to probe power law behaviour is to create a double logarithmic histogram of the frequency distribution. If while doing so we notice a straight line form, this is an indication of the distribution following a power law and the gradient of this line is our scaling parameter; which in the case of on-off intermittency is $-3/2$. We present statistical checks to analyse the strength of our findings after fitting a least squares linear regression line to conclude finally.

Contents

Chapter 1: Introduction

This thesis concentrates on the extreme events within the electricity market, specifically the Nord Pool. The price of electricity is the core index of the market; it possesses a 'spiky' nature; within a few hours, prices can upsurge tenfold. In this thesis, a price spike is formally defined as a state such that the spot price surpasses a set threshold fairly outlying from the standard range of fluctuations. These spikes are usually unanticipated creating extreme volatility; a pronounced characteristic within this market. A particular feature of the Nord Pool is short-lived price spikes that drop back to a mean level within a short period. Prices consequently cluster around the mean level, while those that do not give rise to our study of extreme values.

We aim to identify or reject the hypothesis of our spot price data behaving in terms of on-off intermittency. On-off intermittency is ultimately characterised by two states, on, representing the burst and off, representative of the mean level state. It was observed initially in the system of coupled identical chaotic maps, and further studies saw the power-law distributions of the nearly level state adhere to a rule. "At the onset of intermittent behaviour, the distribution of laminar phases for a large class of random driving cases exhibits a universal asymptotic -3/2 power law." (Heagy, Platt and Hammel, 1994).

In this work, we consider the events of extreme price variations in electricity prices from the Nord Pool market over eight years from 1999. In chapter 2, we provide an overview of the Nord Pool market and the data we will analyse, along with some characteristics we find from this data set. We will also learn about power laws, intermittency and returns. Chapter 3 analyses the properties we draw from the time intervals between extreme events. To capture these extreme events, we set a threshold value. To carry out this study, we use the program R to produce our figures and findings. We finally reach a methodical conclusion in chapter 4.

Chapter 2: Background

2.1 Nord Pool

Nord Pool operates efficient, simple and secure power markets in the Nordics, Baltics, Germany and the UK, (Nord Pool). The Nordic power market was established in 1993, and in 2002 the spot market was fully recognised as a self-governing licensed physical power exchange.

Nord Pool runs as a single market with each region equipped with generators to fulfil that areas demand cost-efficiently. However, if for any reason a region's demand is higher than the supply available, electricity is transported via nearby regions. The majority of the Nord Pool system is operated on the day-ahead market, acquiring great stability between the supply and demand. Nonetheless, incidents can arise between the closing from the previous day and delivery on the following day. Intraday trading has been introduced to stabilise, in real-time, the supply and demand and offset sudden unpredicted fluctuations.

The day-ahead market is an auction where each day is split into 24 hourly spot contracts. All participants submit their bids to Nord Pool the day prior, and aggregated supply and demand curves are established, the system price is the equilibrium price for each hour. Bids are made according to how much each buyer and seller will need and be able to deliver respectively, only after the system price is released does the participant identify the cost of their trade. The participant will obtain an exchange quantity parallel to their bid, and power is exchanged in trade for supply each hour of the next day.

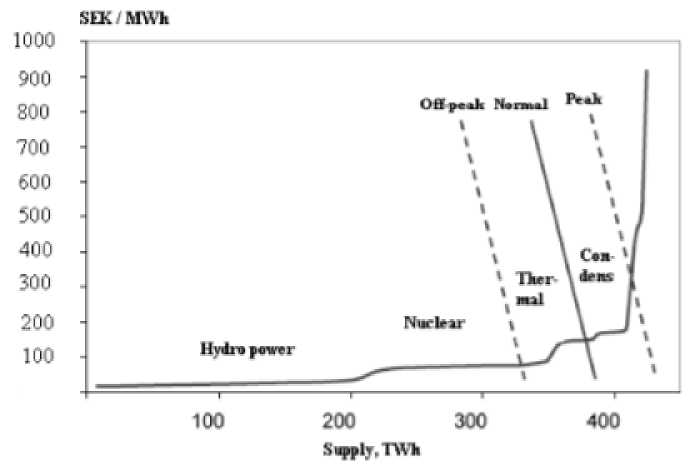


Figure 2.1: The supply and demand levels with the marginal cost for an average year in the Nordic power system. Source: SOU 2004:129, 2004.

Figure 2.1 shows how the Nordic power system uses different production units and marginal costs; hydropower has a low cost of production, along with wind power. Marginal prices rise gradually throughout low demand periods, and renewable energy in the form of hydropower is used; however, it is relatively unpredictable compared to other sources. Consequently, Nord Pool also uses more expensive production units, such as nuclear and thermal production when necessary. Coal and oil condensing are the most costly and resource approximately 10% of the market, typically for peak periods only – condensing is reliable since it generates power rapidly from startup. This increase in demand during peak periods will reflect in the spot prices, causing a rise.

The supply chain is dominated by hydropower - it constitutes for around 60% of power production. Despite the fact that electricity cannot be stored, the hydropower reservoirs hold water, and reservoir levels are an important figure when analysing spot prices. Hydropower electricity production is hugely reliant on the weather; when reservoir levels drop, prices will rise, and accordingly, when reservoir levels are high, prices will decrease. However, the reservoir levels infrequently change a great deal each hour; hence costs of hydropower production is predictably similar to the day before.

2.2 Power Law and Intermittency

Mathematically, a random variable that illustrates a power law distribution is a probability distribution function, for continuous values, defined as

$$p(x) \sim Cx^{-\alpha} \tag{2.1}$$

where α is a constant parameter called the exponent of the power law and C is the normalisation constant. Substantial attention has been drawn towards the power law due to its appearance in numerous natural and manmade phenomena and thus has significant consequences for appropriately modelling and understanding the systems. In practice, it is unusual a studied phenomenon obeys power laws wholly for all values of x, rather the power law applies only for values greater than some minimum, x_{min}, (Clauset, 2007), namely the tail of the distribution follows a power law. Equation 2.1 can be identified as

$$\log p(x) \sim \log C - \alpha \log x \tag{2.2}$$

on a double logarithmic plot; that is to say, a power law distribution follows a straight line on log-log axes. Commonly a least squares linear regression is performed to appropriate the slope of the logarithmic histogram.

The concept of intermittency describes the random transition of two different phases; from laminar, moderately stable, to turbulent, relatively irregular, in a time series, despite the control parameters remaining constant and no presence of considerable external noise. By way of explanation, the chaotic interchanges between phases with large and small volatilities. Intermittency characterises many complex dynamical systems and this phenomenon can be seen in numerous fields; for example, earthquake occurrences [5], the calm of the cyclic solar activity disturbed by short bursts [10] and fluid dynamics [11].

In 1993, Platt, Spiegel and Tresser, first discovered the aperiodic switching known as “on-off intermittency”, where one or more dynamical variables possess a distinct pattern of two states amidst the time series. In the “off” (laminar) state the system stays in the vicinity of some nearly constant value for a long period of time and in the “on” (turbulent) state it chaotically bursts out away from this approximately constant value and returning swiftly. Heagy, Platt and Hammel in 1994 characterised the universal statistical properties of on-off intermittency in systems; they confirmed that, unlike other forms of intermittency, on-off intermittency is parameter bounded and the probability density function of the duration of the laminar phases follows a power law with an exponent equal to $-3/2$.

In 2007 Bottiglieri and Godano reported the time series of earthquake occurrences exhibited on-off intermittency. There were stages of clustered occurrences of earthquakes followed by inactive phases. They concluded that the 'behaviour depends on the value of the threshold', chaotic behaviour was seen for a low threshold value but with a threshold too high the data became too poor to conclude any trend patterns. Reviewing the results they obtained (Figure 2.2) suggests the slopes are fitting with on-off intermittency.

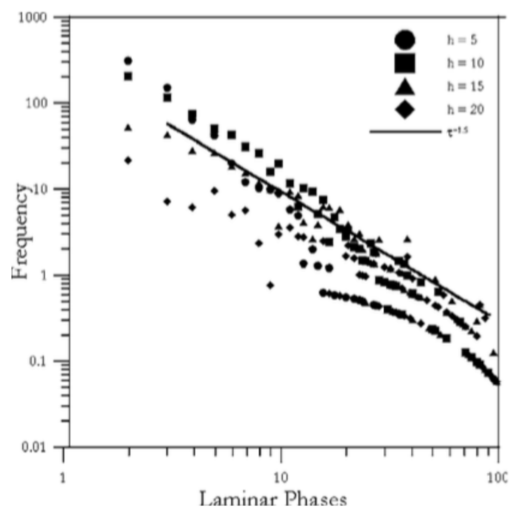


Figure 2.2: Distributions of laminar phases duration of earthquake occurances. Source: M. Bottiglieri and C. Godano. On-off intermittency in earthquake occurrence. Physical Review E 75, 026101 2007.

2.3 Spot Price Analysis

Our spot price analysis concentrates on hourly spot prices in the Nord Pool power market, prices are given in Eur/MWh, and our data set comprises of hourly observations over the period of January 1st 1999 to January 26th 2007, amounting to 70752 data points. Figure 2.3 presents a plot of the raw data points and displays price fluctuations over each hour. At first glance periods of tremendous irregular volatility are observed, and price spikes span the whole length of our data set but generally are shorted lived and display behaviour of returning to an equilibrium price level - mean reversion. The most significant price spike was 238.01 Eur/MWh, compared to the lowest price of 2.3 Eur/MWh, and it is easy to see the immense difference between the greatest and lowest prices.

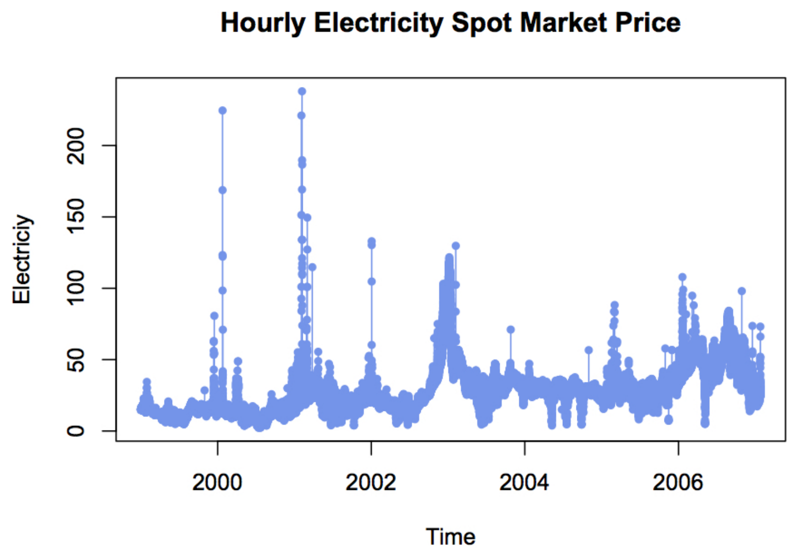


Figure 2.3: Hourly spot prices in Nord Pool power market from 1999 to 2007

2.4 Characteristics

As of currently, there is no efficient way to store vast amounts of electricity; consequently, it has to be utilised as soon as it is produced – the supply and demand have to be steady at all times. Since we cannot store electricity sufficiently, there is little flexibility in running the electricity market without failing the whole network, therefore, operating with a certain capacity is most cost-efficient for the majority of plants. The absence of storability along with the unpredictable demand due to unforeseen events, such as severe unscheduled weather conditions, transmission failures or power plant collapses enhances the instability, unsettling the delicate equilibrium between supply and demand and triggering temporary price spikes. Price spikes “occur due to frictions in demand and/or supply to which electricity producers cannot respond flexible enough" (Huisman, 2007), and a direct consequence, significant price spikes are seen in the spot market. Price spikes are sudden price rises directly shadowed by an abrupt decline and reversion to the mean price. The spiky behaviour can be seen in the graph of the spot price data in Figure 2.3, displaying high volatility and unpredicted short-lived price changes.

Figure 2.3 reveals some preliminary indications of volatility clustering, "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" (Mandelbrot, 1963). As the electricity market reacts to new sudden events and responds with a price increase, these high volatility circumstances last for an extended period following the jolt; a relatively stable period with considerably lower volatility will follow - volatility typically clusters.

2.5 Returns

Now to improve the interpretability of our graph, we will transform our dataset and plot the logarithmic hourly returns to identify the extreme events visually.

As explained earlier, in a short period electricity spot prices can vary remarkably. The return is a widely adopted measure of price fluctuations in various financial markets. A return is a relative change in a variable expressed as a percentage, which lets us compare the variable over altered time periods. We will now solely study the return data as opposed to the price data since the returns reveal more motivating statistical properties. They are not based upon the time scale of data, rather, relational to the prices immediately preceding them, yielding a more stable means to study behaviour over a length of time.

Simple return is defined as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

where P_t is the spot price at time t. Simple return is the relative variation in net asset value from time $t-1$ to t.

The log return is defined as

$$r_t = log(\frac{P_t}{P_{t-1}}) = log(P_t) - log(P_{t-1})$$

where $log(P_t)$ is the logarithmic price at time t and note that a log return uses the log function with the natural base.

Log returns are far superior to simple returns since they have the immediate advantage of being time additive. For example, using the case of returns from t to $t+n$ the log return $r_{t,t+n}$ would simply be the sum of individual log returns within these times,

$$r_{t,t+n} = \sum_{i=1}^{n} r_{t+i,t+i-1}$$

Conversely, simple multi-period returns are the product of individual log returns from t to $t+n$, which mathematically is not as convenient as using logs. Log returns are assumed to follow a normal distribution, and again, the sum of continual samples from a normal distribution is also normally distributed, making the use of log returns more advantageous.

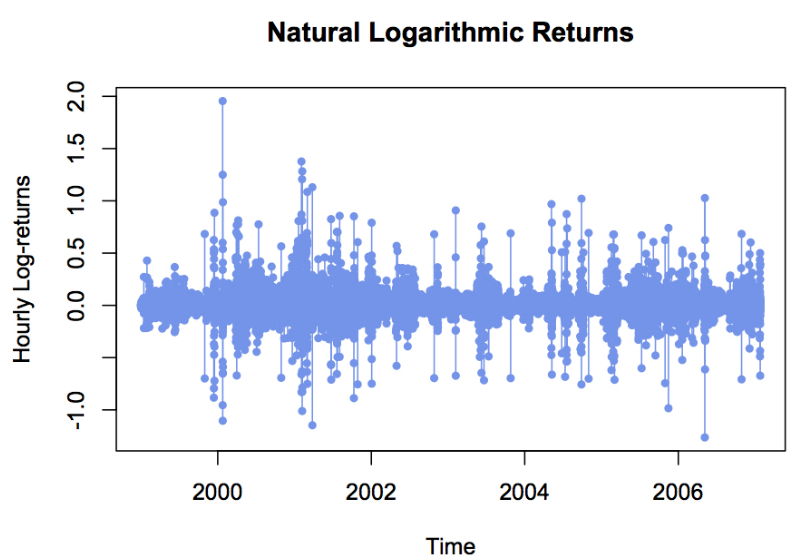
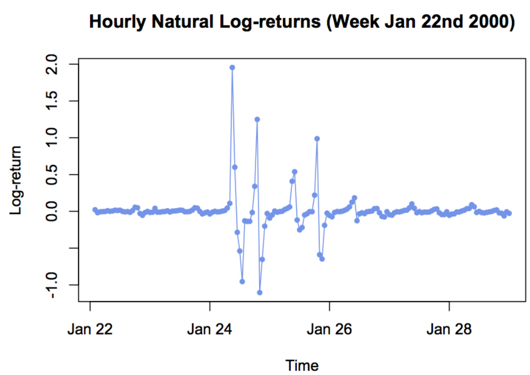


Figure 2.4: Hourly natural logarithmic returns in Nord Pool power market from 1999 to 2007

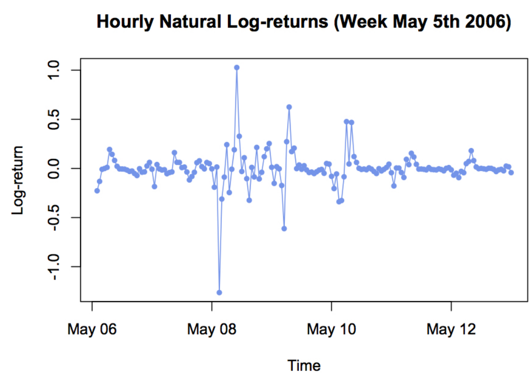
Figure 2.4 presents hourly natural logarithmic returns based on the Nord Pool data versus time; we see that the method presents slightly fewer variance fluctuations throughout time and so is more stable, however, is still very volatile. The volatility is most extensive at the beginning of 2000, with greatest hourly log return being 1.954708 and the smallest value -1.263635, with a massive difference of 3.218343. The time series indicates on-off intermittency - large stretches of almost regular prices disrupted by large outbursts.

Chapter 3: Analysis

In Figure 3.1 we have a more detailed look at the hourly returns across a one-week duration containing the occurrences of the highest and lowest values of hourly log returns. Figure 3.1a plots the week beginning January 22nd 2000 and Figure 3.1b plots the week beginning May 5th 2006.



(a) Hourly log returns for the week beggining Jan 22nd 2000



(b) Hourly log returns for the week beggining May 5th 2006

Figure 3.1: Hourly log returns across a one-week duration

Electricity demand is generally incredibly inelastic; however, price sensitivity is a consequence of supply elasticity. The occurrence of extreme events means the baseload generation from nuclear and hydropower, produced at a low cost, is insufficient. Other rare unforeseen circumstances may occur, having a substantial impact on the demand and supply balance. The leading cause of price increases is the temperature in the Nordic countries; extreme cold temperatures require power demand for heating.

3.1 Time Series Plots

In this section, we will concentrate on the extreme points following a power law and aiming to identify on-off intermittency. Detecting the behaviour of a power-law in both man-made and natural systems may be a complex task. The standard approach asserts that a histogram concerning a quantity following a power-law distribution when plotted on logarithmic scales has the appearance of a straight line. We must first adjust our data and acquire from it the data points we will use in our histograms as follows.

As illustrated in Figure 2.4, the average of the natural log return is almost zero, however it is still extremely volatile. We produce a plot of the deviations from the mean by calculating the mean of the return and discounting for that value so that the resulting time series has mean zero, (Figure 3.2). A deviation is a measure of the difference between the observed value and the mean value, displaying the dispersion of the data. Visually our plot looks indistinguishable to the original series, as the mean is so small.

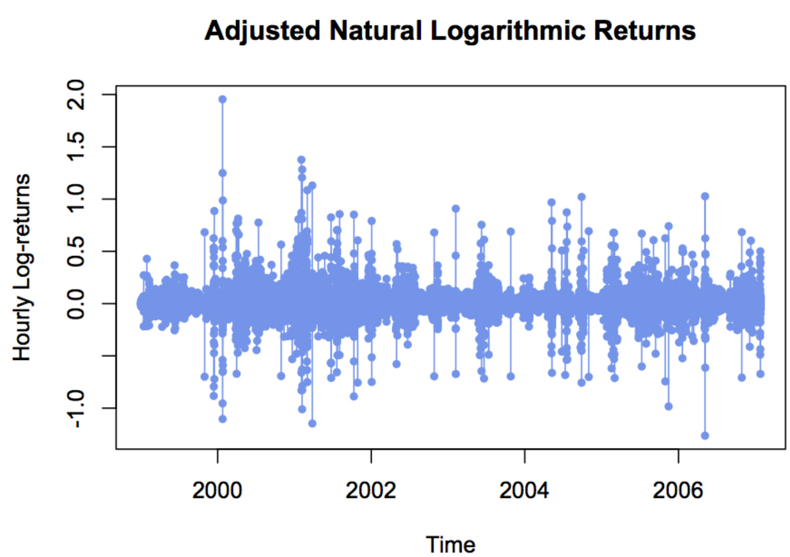


Figure 3.2: Hourly natural logarithmic returns with the mean discounted

Now there are "positive" and "negative" extreme events and one could distinguish between both, but to make our problem simpler we will treat both on the same level.

The absolute value plot of our adjusted natural logarithmic returns time series takes all the negative values and produces their respective positive values, (Figure 3.3).

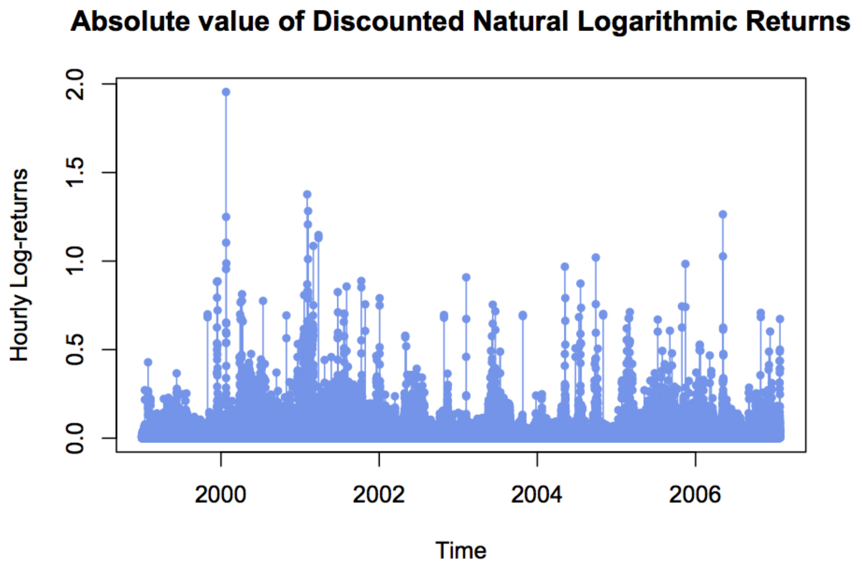


Figure 3.3: Absolute discounted hourly natural logarithmic returns

As outlined previously, we now need to identify extreme events, i.e. discounted returns which exceed a certain threshold. We have to select a suitable threshold value, such that if absolute discounted return points exceed this value, we define them to be an extreme event. That is, an event is called extreme if the return deviates from the mean by a certain amount, the threshold value. If our threshold were too small, it would be likely all our events become extreme; which would be nonsense, and if it were too large, there would be almost no extreme events. Essentially there is a whole range of threshold values, but to get started we chose the value 0.5.

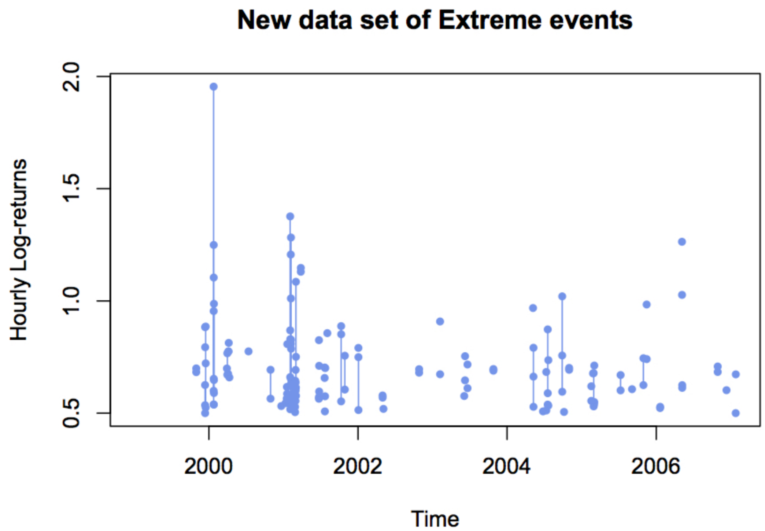


Figure 3.4: Data points from our discounted hourly natural logarithmic returns which exceed the 0.5 threshold

The introduction of the threshold value allows us to ignore all events with time indices less than 0.5, leaving us with a new data set, graphed in Figure 3.4.

3.2 Histogram of Time Differences

We want to investigate the distribution of time spans between extreme events, namely, between massive increases or decreases of the market prices. Now to do so, we compute the difference between subsequent entries and then construct a histogram of these times between extreme events. Bins with zero observations are excluded because $log(0)$ is undefined, and later on we will need to use the logarithmic scale. Also, we do not need to worry about small differences, for example, a difference of one or two may not indicate the difference between two different extreme events as the data points could belong to the same event. As for evaluating the distribution, we can then look past the small values, say one and two.

From studying our histogram in Figure 3.5 we identify that most differences are in the first bin and in all the other bins there are a lot less; our histogram is extremely right skewed. In Figure 3.6 we plot the time difference of all log returns above our threshold but now on a log-log scale. Since we are aiming to identifying a power law, $p(T) \sim CT^{-\alpha}$, a double logarithmic plot, $\log p(T) \sim -\alpha \log T + \log C$, will yield results which are easier to visualise, i.e. the histogram should look like a line when $log(P)$ is plotted versus $log(T)$.

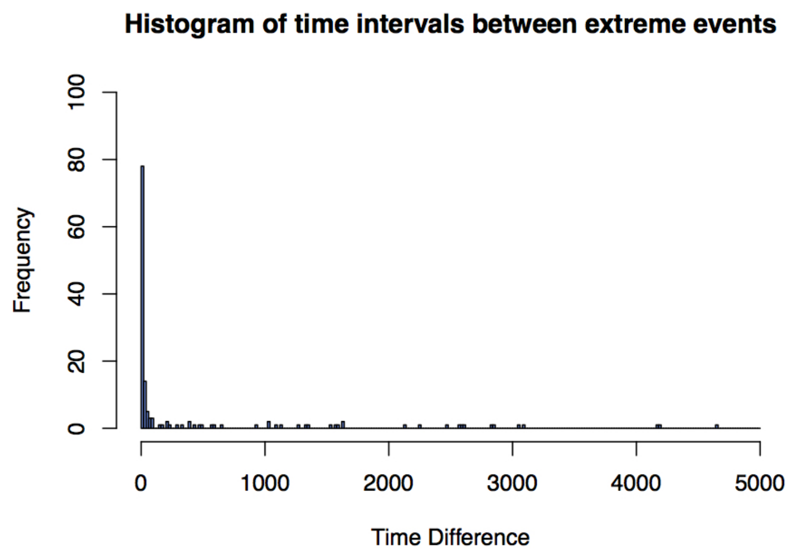


Figure 3.5: Histogram of the time difference of logarithmic returns above the threshold of 0.5

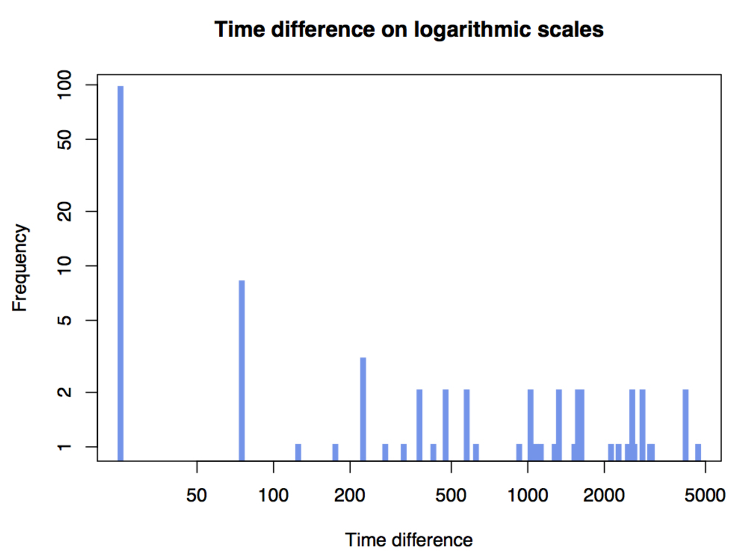


Figure 3.6: Histogram of the time difference of logarithmic returns above the threshold of 0.5 on logarithmic scales

3.2.1 Linear Regression

As already discussed, a straight line on the double logarithmic plot suggests the data follows a power law. Our next task is to fit such a straight line and estimate the exponent parameter α. In fact, we are aiming to identify on-off intermittency characterised by the $-3/2$ law. Using the function 2.1 the probability distribution function will be $p(x) \sim Cx^{-3/2}$, and since we have the logarithmic plot, taking the logarithm of both sides we learn that the power law obeys the relation $\log p(x) \sim \log C - 3/2 \log x$, the equation of a straight line.

If we can spot that our points on the histogram lie on a straight line we can draw this line in and continue it to the $log(y)$ axis to find the point at which it crosses; this value will correspond to $log(C)$, where C is our normalisation constant. Likewise, the gradient of the line is interpreted as the estimate of α, in our power law function. To fit this straight line, we will use the algorithm `lm()` in R, which fits the function to the linear form by the least squares linear regression.

A linear regression is a model that explores the relationship between two variables and establishes if it is statistically meaningful. The two variables play very different roles; the dependent variable is the one whose value we want to explain, and the independent variable helps explain the variance in the dependent variable. In our data, the counts is the dependent variable, and the time difference is the independent variable.

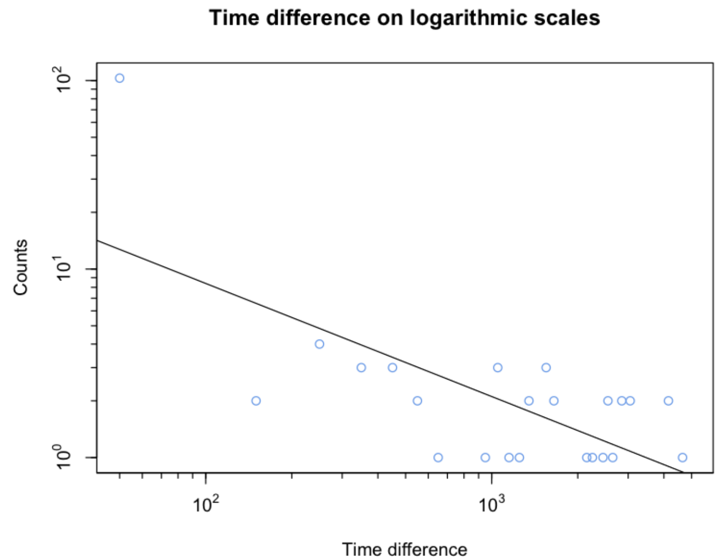


Figure 3.7: Histogram with a fitted regression line

Least Squares Estimation

Essentially, we have a set of observations and we need to estimate the values of the coefficients α and β, using the data. The least squares principle yields these coefficients by minimizing the sum of the squared residuals. Each residual is given by $\hat{\varepsilon}_t = y_t - \alpha - \beta x_t$ for all parameter values α and β. In other words, minimise

$$\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

The estimation takes the difference between the data and the predicted line of best fit, calculates the squared error, and the line continually shifts until it is the least possible distance from each point of the data.

A scatter plot of time differences against density is shown in Figure 3.7, together with the computed regression line $\hat{y}_t = 2.1237 - 0.6001x_t$, ($\hat{y}$ indicates that the value of y is a prediction). The fitted line posses a negative gradient, illustrating further the negative relationship between time differences and the counts. The slope of the line is -0.6001, which is not close at all to $-3/2$; however, we will not disregard the on-off intermittency hypothesis, we will use different thresholds and study further the ways to test the significance of our coefficient value.

3.2.2 Analysis with Different Thresholds

Initially, we used a high threshold of 0.5, and it is noteworthy that the data did not correspond closely to the behaviour of on-off intermittency. Using such a high threshold to define an extreme event lead us to discarding a considerable number of points from our data set of returns (Figure 3.3), motivating us to test a lower threshold. If the threshold is too low, the period of the laminar phases turn out to be too brief, and the occurrence is inadequate to be classified as intermittency. Although as we have just seen if, the threshold is too high, there is a loss of scaling invariance, which is needed since we need our results to be somewhat independent of the threshold value. We followed the same steps as previously, but now using the threshold 0.2 and arrived at the histogram graphed in Figure 3.8, with a gradient -1.470, very close to -1.5, illustrated in red, the PDF for on-off intermittency. This closeness may imply that our data follows on-off intermittency behaviour; however, we will now look at the linear regression and decide whether it is a good enough fit to claim the hypothesis.

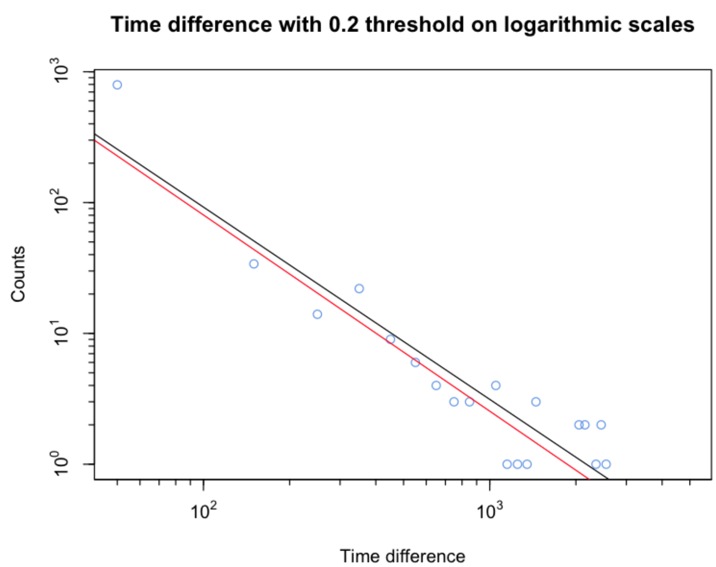


Figure 3.8: Histogram of time differences above 0.2 threshold with a fitted regression line, plotted on logarithmic scales. The red line corresponds to the power law $p(x) \sim Cx^{-3/2}$

.

3.2.3 Linear Regression Analysis

In R, the command `summary()` provides a superb report of our regression model. We will now look at the linear model fit and explain the components of the model output.

```
> summary(reg2)

Call:
lm(formula = logcounts ~ log10(hi$mids))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.40615 -0.18843 -0.05087  0.19791  0.49206 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)      4.9051     0.4071   12.05 9.44e-10 ***
log10(hi$mids)  -1.4699     0.1387  -10.60 6.59e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2674 on 17 degrees of freedom
  (31 observations deleted due to missingness)
Multiple R-squared:  0.8685,Adjusted R-squared:  0.8607 
F-statistic: 112.3 on 1 and 17 DF,  p-value: 6.594e-09
```

Figure 3.9: A screenshot of the the command `summary()` output in R corresponding to our linear regression model

Residuals

A reasonable method to test the degree of suitability of the model is to analyse the residuals; the differences between the values the model predicted and the observed values. Essentially, a residual quantifies how far a data point is from the regression line. The main idea behind our analysis is we want a symmetrical distribution across the residuals and for the sum to be as low as possible, ideally near zero. In our model, the dispersal of residuals is remarkably symmetrical around the mean value zero. Taking this idea slightly further, we will plot the residuals on a normal probability plot - a means toward comparing our linear regression model differences against the normal distribution. From studying Figure 3.10, our normal probability plot of the residuals, it appears the relationship is approximately linear since the residuals fall very close to the straight line. We can conclude that the differences are normally distributed.

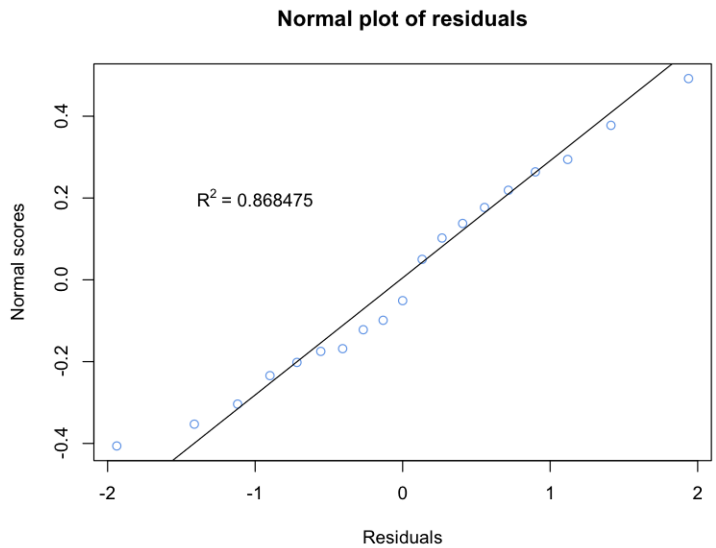


Figure 3.10: The normal probability plot of the residuals

The statistic R-squared determines to what extent the model fits with the actual data. It is understood as a ratio of variation in the response variable (counts) described via variation in the predictor variable (time differences). The R-squared statistic stands within the range 0 and 1; a result close to 1 indicates the model closely tracks the data under consideration, whilst a score close to 0 signifies low correlation. The R-squared we have is 0.8685, meaning the time differences can justify approximately 87% of the variance seen in the counts – this is a relatively strong R-squared. However, the task of defining an appropriate level of R-squared to assert that the data fits appropriately is difficult. Since we are looking at bivariate data, we do not need to worry about the adjusted R-squared.

Coefficients

The following section in the model output analyses any variables in the model and are listed under the coefficients, in our case, we only have one, along with the intercept. The coefficients correspond to the following model

$$\text{log density} = \textit{intercept} + \text{log time difference} * \textit{slope}$$

The two coefficients are the unknown constants that characterise the intercept and slope terms in the linear regression model – fitting the line such that the data points are as close to it as possible. The estimated coefficients, gives rise to log density $=$ $4.9051 +$ log time difference $* -1.4699$. The intercept measure is made exclusive of any variable; it is the intercept value when the other variable is 0. The estimates have some standard errors attached to them; the standard errors are the variability we would expect if we did repeated sampling; this captures sampling variability and also shows practical significance.

The t-value is simply a ratio of the estimate divided by their standard error, enumerating how big the estimate is relative to its standard error. Associated to the t-value is a p-value, found in the model since the t-value falls on a distribution. The p-value reports the probability of observing a value of t or larger, telling us how statistically significant the estimates are. A low p-value implies it is unlikely a relationship between the log density and log time difference is due to chance – normally a p-value of 0.05 or less is an appropriate conclusion point. Both of our p-values are well below 0.05; thus, the estimates are both statistically significant. The significance, symbolised by three asterisks, as a quick interpretation, suggest highly significant p-values for this model.

3.2.4 Bin Width Analysis

At the outset, we started with a threshold of 0.5, and this produced a histogram such that we decided it did not follow on-off intermittency behaviour, and then we used a 0.2 threshold and concluded since the linear regression was a good fit, this did. Looking back to Figures 3.7 and 3.8 we used a bin width of 100 to produce the histograms. Using the argument `breaks` in the command `hist` in R, to adjust the bin width, we will check how changing of the bin width affects the outcomes. When we lowered the threshold of the extreme event to 0.2 we were left with 908 data points in comparison to only 145 when the threshold was 0.5 – recall a data point is counted to be in the “off” state when its value is lower than the threshold. We found that changing the bin width has a significant effect on how the histogram will look, as the data gets grouped differently. We obsereved that for the histogram with threshold 0.5 we can get a slope somewhat close to -1.5; using the bin width 600 lead us to a slope with gradient -1.516. However, a bin width of 600 for only 145 data points is relatively large in comparison; we produce poor statistics. Evaluating the distributions while increasing the bin width reduces the information about the variability of the distribution since the data is grouped into fewer bins. When we have a look at a third threshold of 0.3, grouping using bin width 100 produces a slope with gradient -1.108; however, we acquire its linearity with slope -1.54, again similar to -1.5, if we increase the bin width to 225 – indicating on-off intermittency holds once again for this new threshold and bin width. When the bin width assumes higher values, we can observe on-off intermittency for higher thresholds, though lower thresholds require lower values for the bin width.

Chapter 4: Conclusions

This project shows that extreme values of electricity spot prices can be characterised using on-off intermittency. We used hourly electricity spot prices from the Nord Pool market, from 1999 to 2007 - over 70,000 spot prices. This data exhibited volatility leading to short-lived price spikes which returned to a somewhat constant state, a very significant trait of on-off intermittency. We investigated the distribution of periods between extreme events, i.e., between massive increases and decreases of the market prices. We studied the histograms of the data on double logarithmic plots, and these plots seemed to be linear with a gradient near to -1.5, the consistency of the data with on-off intermittency behaviour was concluded.

To identify on-off intermittency, the scaling should be, to a great extent, independent of the threshold (i.e. it should appear for a range of values, and the slope should only differ a little). Changing the threshold level altered the results quantifiably; nonetheless, we accordingly adjusted the bin width. For low values of the threshold, intermittency was seen using a reasonable bin width; however, for a greater threshold value, the statistics became too poor. Since we observed a gradient value close to -1.5 for multiple thresholds after adjustment, this indicates that on-off intermittency prevails – the scaling became reasonably stable. The need for changing bin widths emphasises the complexity of characterising genuine on-off intermittency power-law behaviour. We also analysed our linear regression and very quickly saw our regression was a good fit. Ultimately, our results suggest that extreme values of electricity spot prices can be deemed a model of on-off intermittency.

Chapter 5: Appendix: R Code

Here are screenshots of the commands in R used to generate each graph included in this thesis.

Figure 2.3: Hourly spot prices in Nord Pool power market from 1999 to 2007.

```
> setwd('/users/Elliz/Documents')
>    data <- read.table("BEUR.txt")
>    datatimeseries <- ts(data)
> vs<-data[,1]
> StartTime <- as.POSIXlt("1999-01-01 00:00:00")
> EndTime <- as.POSIXlt("2007-01-27 00:00:00")
> Time <-difftime(EndTime, StartTime, units= "hours")
> x <- strptime("1999-01-01 00:00:00", "%Y-%m-%d %H:%M:%S")+3600*1:time
> plot(x, data[,1], type= "o", pch=20, xlab="Time", ylab="Electriciy", main ="Hourly Electricity
Spot Market Price", col=" cornflowerblue")
```

Figure 2.4: Hourly natural logarithmic returns in Nord Pool power market from 1999 to 2007.

```
> rs<-diff(log(vs), lag=1)
> plot(x[2:70752], rs, type= "o", pch=20, xlab="Time", ylab="Hourly Log-returns",
col="cornflowerblue", main="Natural Logarithmic Returns")
```

Figure 3.1(a): Hourly log returns across a one-week duration for the week beginning Jan 22nd 2000.

```
> date1<- as.Date("2000-01-22","%Y-%m-%d")
> date2<- as.Date("2000-01-29","%Y-%m-%d")
> subdf<- difftime(date2, date1, units="hours")
> print(subdf)
Time difference of 168 hours
>
> StartTime<- as.POSIXlt(date1)
> Plotx<- StartTime+3600*2:168
> plot(Plotx, rs[9264:9430], type="o", pch=20, xlab="Time", ylab="Log-return", main="Hourly
Natural Log-returns (Week Jan 22nd 2000)", col=" cornflowerblue")
```

Figure 3.1(b): Hourly log returns across a one-week duration for the week beginning May 5th 2006.

```
> date3<- as.Date("2006-05-06","%Y-%m-%d")
> date4<- as.Date("2006-05-13","%Y-%m-%d")
>  subdf<- difftime(date4, date3, units="hours")
>  print(subdf)
Time difference of 168 hours
>
> startTime<- as.POSIXlt(date3)
> Plotx<- startTime+3600*2:168
> plot(Plotx, rs[64344:64510], type="o", pch=20, xlab="Time", ylab="Log-return", main="Hourly
Natural Log-returns (Week May 5th 2006)", col=" cornflowerblue")
```

Figure 3.2: Hourly natural logarithmic returns with the mean discounted.

```
> mean(rs)
[1] 6.692666e-06
>
> Dis <- rs-mean(rs)
>
> plot(x[2:70752], Dis, type= "o", pch=20, xlab="Time", ylab="Hourly Log-returns",
col="cornflowerblue", main="Adjusted Natural Logarithmic Returns")
```

Figure 3.3: Absolute discounted hourly natural logarithmic returns.

```
> Abs <- abs(Dis)
> plot(x[2:70752], Abs, type= "o", pch=20, xlab="Time", ylab="Hourly Log-returns",
col="cornflowerblue", main="Absolute value of Discounted Natural Logarithmic Returns")
```

Figure 3.4: Data points from our discounted hourly natural logarithmic returns which exceed the 0.5 threshold.

```
> length(An[An>0.5])
[1] 145
> Bn <- An
> Bn[An < 0.5] <- NA
> plot(x[2:70752], Bn, type= "o", pch=20, xlab="Time", ylab="Hourly Log-returns", col="
cornflowerblue", main="New data set of Extreme events")
```

Figure 3.5: Histogram of the time difference of logarithmic returns above the threshold of 0.5.

```
> Cn<- na.omit(data.frame(Bn))
> Dn <- cbind(DIFF1 = rownames(Cn), Cn)
>    rownames(Cn) <- 1:nrow(Cn)
>
> index<- as.numeric(as.character(Dn[, 1]))
> tail(index, -1) - head(index, -1)
  [1]    1 1036    1    2    5    4   60    1    2
 [10]  933    1    2    1    6    1    1   13    9
 [19]    1    1 1524   49    1   95   23   24   24
 [28]   73 2254 2588    1 1263  494  144   23   24
 [37]   24  336    1    3    7    1   59    1    1
 [46]    4    4    4   11    1    3    7    1  468
 [55]    1   23   24    1    1    9    2   35    1
 [64]    3    7  566    1 2129    1    4   21    3
 [73]  644    4   26   21   24  240 1626    1    4
 [82]  429    1 1597    3    1 2837    6   97 4194
 [91]    1 2462    2 2842   92    5  285    4 3043
[100]    1 4658   49    2    1 1124  383    2  163
[109]    7    1   66    2 1627    5    1  214  598
[118]    1 2606    2  205   72    1   47   23    3
[127] 3092    1 1345 1338    1  381    5 1574   12
[136] 2562    7   19    2 4173    1 1034 1083   10
> Data2<- tail(index, -1) - head(index, -1)
>
> hist(Data2, xlim=c(0,5000), ylim=c(0,100), breaks=seq(from=0, to=5000, by=20), xlab= "Time
Difference", main="Histogram of time intervals between extreme events", col="cornflowerblue")
```

Figure 3.6: Histogram of the time difference of logarithmic returns above the threshold of 0.5 on logarithmic scales.

```
> h <- hist(Data2, plot=F, breaks=100)
>   plot(h$mids, h$counts, log="xy", pch=20, col="cornflowerblue",
+            main="Time difference on logarithmic scales",
+          xlab="Time difference", ylab="Frequency", type="h", lwd=5, lend=2)
```

Figure 3.7: Histogram with a fitted regression line.

```
> hi<- hist(Data2, xlim=c(0,5000), ylim=c(0,100), breaks=seq(from=0, to=5000, by=100), xlab=
"Time Difference", main="Histogram of time intervals between extreme events",
col="cornflowerblue")
>
> log10.axis <- function(side, at, ...)
+      {at.minor <- log10(outer(1:9, 10^(min(at):max(at))))
+      lab <- sapply(at, function(i) as.expression(bquote(10^ .(i))))
+      axis(side=side, at=at.minor, labels=NA, tcl=par("tcl")*0.5, ...)
+      axis(side=side, at=at, labels=lab, ...)}
>
>
> plot(log10(hi$mids),log10(hi$counts),  xaxt="n", yaxt="n", main="Time difference on
logarithmic scales", xlab="Time difference", ylab="Counts",  col=" cornflowerblue")
> log10.axis(1, at=seq(0, 4, 1))
> log10.axis(2, at=seq(0, 4, 1))
>
>
> yy<- log10(hi$counts)
>  yy[mapply(is.infinite, yy)] <- NA
> reg<- lm(yy ~ log10(hi$mids))
> abline(reg)
>  reg

Call:
lm(formula = yy ~ log10(hi$mids))

Coefficients:
  (Intercept)  log10(hi$mids)
      2.1237         -0.6001
```

Figure 3.8: Histogram of time differences above 0.2 threshold with a fitted regression line, plotted on logarithmic scales.

```
> Bn2 <- An
>  Bn2[An < 0.2] <- NA
>
>     Cn2 <- na.omit(data.frame(Bn2))
> Dn2 <- cbind(DIFF1 = rownames(Cn2), Cn2)
>        rownames(Cn2) <- 1:nrow(Cn2)
>
>    index<- as.numeric(as.character(Dn2[, 1]))
>
>    Data2b<- tail(index, -1) - head(index, -1)
>
> hi<- hist(Data2b, xlim=c(0,5000), ylim=c(0,100), breaks=seq(from=0, to=5000, by=100),
xlab= "Time Difference", main="Histogram of time intervals between extreme events",
col="cornflowerblue")
>
>  log10.axis <- function(side, at, ...)
+       {at.minor <- log10(outer(1:9, 10^(min(at):max(at))))
+       lab <- sapply(at, function(i) as.expression(bquote(10^ .(i))))
+       axis(side=side, at=at.minor, labels=NA, tcl=par("tcl")*0.5, ...)
+       axis(side=side, at=at, labels=lab, ...)}
>
>
>  plot(log10(hi$mids),log10(hi$counts),  xaxt="n", yaxt="n", main="Time difference with 0.2
threshold on logarithmic scales", xlab="Time difference", ylab="Counts",  col=" cornflowerblue")
>  log10.axis(1, at=seq(0, 4, 1))
>  log10.axis(2, at=seq(0, 4, 1))
>
>  logcounts<- log10(hi$counts)
>   logcounts[mapply(is.infinite, logcounts)] <- NA
>  reg2<- lm(logcounts ~ log10(hi$mids))
>  abline(reg2)
> abline(        4.905          , -3/2, col="red")
>
>   reg2

Call:
lm(formula = logcounts ~ log10(hi$mids))

Coefficients:
  (Intercept)  log10(hi$mids)
       4.905          -1.470
```

Figure 3.10: The normal probability plot of the residuals.

```
> resid(reg2) #List of residuals
          1           2           3           4
 0.49206105 -0.17495672 -0.23421060  0.17687865
          5           6           7           8
-0.05086932 -0.09885793 -0.16830646 -0.20189357
          9          11          12          13
-0.12199286  0.13783978 -0.40614634 -0.35291782
         14          15          21          22
-0.30378801  0.21895068  0.26391263  0.29431706
         24          25          26
 0.05006869  0.37770139  0.10220970
>       plot(density(resid(reg2)))
>  qqnorm(resid(reg2), xlab="Residuals", ylab="Normal scores", main="Normal plot of residuals",
col="cornflowerblue")
>  qqline(resid(reg2))
> summary(reg2)$r.squared
[1] 0.868475
> text(-1, 0.2, expression(R^2 ~ "= 0.868475"))
```

Bibliography

[1] A. Clauset, C.R. Shalizi and M.E.J. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4): 661–703, 2009

[2] J. F. Heagy. Characterization of on-off intermittency. *Physical Review E*, 49(2): 1140-1150, 1994

[3] R. Huisman. The influence of temperature on spike probability in day-ahead power prices. *Energy Economics* 30(5): 2697 – 2704, 2007

[4] B.B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36(4): 392–417, 1963

[5] M. Bottiglieri and C. Godano. On-off intermittency in earthquake occurrence. *Physical Review E*, 75(2), 2007

[6] N. Platt. On-off intermittency: a mechanism for bursting. *Physical Review Letters*, 7(3), 1993

[7] O. Knapik. Modeling and forecasting electricity price jumps in the Nord Pool power market. *Center for Research in Econometric Analysis of Time Series*, Denmark, 2017

[8] Y. Virkar and A. Clauset. Power-law disributions in binned empirical data. *The Annals of Applied Statistics*, 8(1): 89-119, 2014

[9] A. Čenys, A.N. Anagnostopoulos and G.L. Bleris. Distribution of laminar lengths for noisy on-off intermittency. *Physics Letters A*, 224(6): 346-352, 1997

[10] N. Platt, E.A. Spiegel, and C. Tresser. The intermittent solar cycle. *Geophysical and Astrophysical Fluid Dynamics*, 73(1): 147–161, 1993

[11] M. Núñez. Rigorous bounds on intermittent bursts for turbulent flows. *Physica D: Nonlinear Phenomena*, 176(3-4): 237–241, 2003

[12] T.M. Christensen, A.S. Hurn and K.A. Lindsay. Forecasting spikes in electricity prices. *International Journal of Forecasting*, 28(2): 400–411, 2012

[13] H. Erzgraber, F. Strozzi, J-M Zaldivar, H Touchette, E. Gutierrez and D.K. Aarrowsmith. Time series analysis and long range correlations of Nordic spot electricity market data. *Physica A*, 387(26): 6567-6574, 2008

[14] Europex. Nord Pool - Europex. 2020. [online] Available at: <https://www.europex.org/members/nord-pool/> [Accessed 7 September 2020]

[15] Otexts.com. Evaluating The Regression Model | Forecasting: Principles And Practice. 2020. [online] Available at: <https://otexts.com/fpp2/regression-evaluation.html> [Accessed 6 September 2020]

[16] Feliperego.github.io. Quick Guide: Interpreting Simple Linear Model Output In R. 2020. [online] Available at: <https://feliperego.github.io/blog/2015/10/23/Interpreting-Model-Output-In-R> [Accessed 9 September 2020].

[17] Filedn.com. 2020. [online] Available at: <https://filedn.com/ljdBas5OJsrLJOq6KhtBYC4/forarbeten/sou/2004/sou-2004-129.pdf> [Accessed 2 September 2020].

[18] Stockholm School of Economics. Did hydropower generators in Nord Pool exercise market power during 2006? – A simulation analysis of hydro production in Norway and Sweden. Master's Thesis in International Economics.