Mathematics and Statistics BSc Third-Year Project MTH6138， 2020/21

On-off Intermittency in Electricity Spot Price

Chunxi, Zheng

Student ID: 181089449

Supervisor: Dr. Wolfram Just

A thesis presented for the degree of Bachelor of Science with Honors in *Mathematics and Statistics*

School of Mathematical Sciences

Queen Mary University of London

Declaration of original work

This Declaration is made on 6th May 2021.

Student's Declaration: I, Chunxi Zheng, 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 any other sources except where due reference or acknowledgment is made explicitly in the text, nor has any part been written for me by another person.

The referenced text has been flagged by:

1. Using italic fonts, **and**

2. Using quotation marks "...", **and**

3. Explicitly mentioning the source in the text.

Acknowledgment

My deepest gratitude goes foremost to my supervisor, Dr. Wolfram Just, for his invaluable advice on the research project. Without his careful guidance and illuminating instruction during our meetings, this project could not have progressed at each step and finally reached its present form.

I would also like to thank my academic adviser Professor Thomas Prellberg, for providing consistent encouragement to me during my time at the School of Mathematical Sciences.

Abstract

The objective of this thesis is to investigate whether the daily electricity spot price exhibits on-off intermittency characteristic. The preliminary indication of on-off intermittency behavior in a time series is the appearance of two distinctive patterns: “off” state presents a stable quantity within a long duration. On the other hand, the "on" state presents a random and temporary burst and returns to the constant. Furthermore, a previous study found that on-off intermittency holds a universal property: the time differences of inactive periods between bursts follow a power-law distribution with the scaling parameter α equals 3/2.

In this thesis, we analyzed the time series of the logarithmic returns of the Nord Pool power market's daily spot price across eight years. First, we set and vary threshold values for extreme values and use the method of least squared to fit the corresponding straight line on the double logarithmic histogram of the frequency distribution of time difference between extreme events – as power-law presents as a straight line in the double logarithmic plot. So, the gradient for the fitted linear regression line for our data should be approximately -3/2. Then we will check the statistical significance of the fitted linear regression line. We finally conclude that spot price appears a substantial similarity with on-off intermittency when a suitable threshold value for extreme events with an appropriate bin width is chosen.

Contents

1 Introduction

The price dynamics of the electricity market are specific and unique, for electricity is seen as a unique commodity for its non-storable nature and requires a constant balance between supply and demand [1]. Factors heavily influencing the production and consumption of electricity are both volatile and unforeseen. As a result, the time series of the electricity spot price exists unlooked-for short-lived price spikes (Weron, 2014).

Our main objective is to explore whether our electricity spot price data comply with on-off intermittency. We can observe on-off intermittency behavior in many chaotic dynamical systems. The feature of on-off intermittency is that there is aperiodic switching from two typical states: static mean level state， known as “off” (laminar) state; unanticipated deviating away and reverting to the “off” state, known as “on” state. In 1994, on-off intermittency was found to obey a rule: "For a large class of random driving cases, the distribution of laminar phases can be obtained from a price expression: it is proven to follow a power-law distribution with universal asymptotic -3/2 [2](Heagy et al., 1994).”

The structure of this thesis is as follows. Section 2 introduces the Nordic power market and the characteristics of electricity prices. We illustrate the dataset used for the thesis: the daily spot price in the Nord Pool over eight years starting from 1999. Section 3 will apply the review of knowledge about returns, power-law, and on-off intermittency. We will then provide an in-depth statistical analysis of the time intervals between extreme events by setting different threshold values and varying bin widths in Section 4. Finally, in Section 5, we conclude all our findings.

2 The Nord Pool

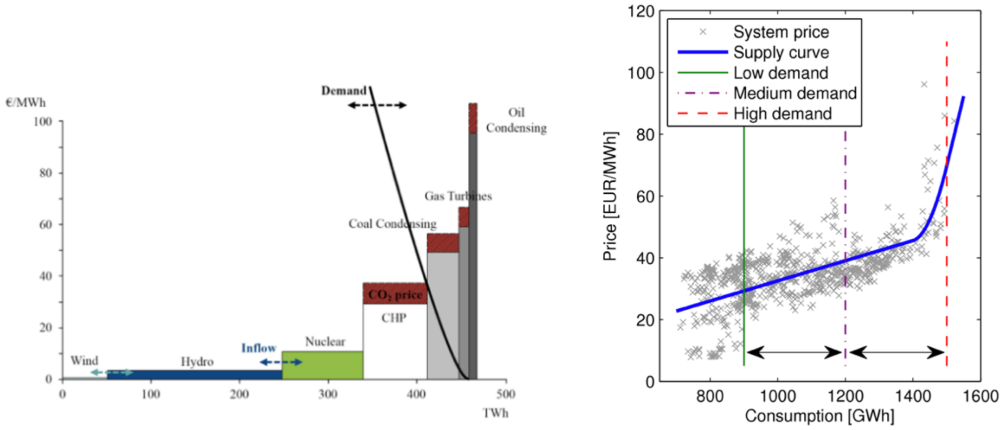
2.1 The Nord Pool Market

The Nordic electricity market, known as the Nord pool, is the lead power market in Europe [3]. It was established in 1993. Until 2005, the power market of Sweden, Finland, the eastern part of Denmark, and some German's region subsequently merged in the Nord Pool. The physical electricity market, Elspot, which our thesis will mainly focus on, was separated from the derivatives market and renamed Nord Pool Spot in 2002 [7]. Nearly 80% of the Nord Pool system is operated on this day-ahead market.

In the Elspot market, all participants trade power contracts for one day-ahead physical delivery. There is an auction for all buyers and sellers to submit their bids for purchasing or selling the hourly power contracts covering 24 hours of the next day [3]. Every day at 12:00 CET, the deadline for final submissions, the (aggregate) supply and (aggregate) demand curves for all buy and sell orders in each time zone are generated, and the system price, also regarded as equilibrium hourly spot price, is determined by the intersection point of the two curves [4].

Though the day-ahead market secures the baseline for planning the balances in the next 24-hour period, unpredictable events may happen when the auction is closed for bidding, causing considerable fluctuations in spot prices. In this case, the Nord Pool also operates an intraday trading system-Elsa. All participants can trade on Elbas around one hour before delivery, enabling them to act on changes and opportunities later in the day. Thus, in real-time, the unforeseen price fluctuations caused by changes in supply and demand can be stabilized by intraday power trading.

A temporary imbalance of the production and the consumption would reflect in the corresponding spot prices.



Left: Figure 2.1 Power production price curve in the Nordic Pool (Source: http://www.nordpoolspot.com) Right: Nordic consumption – price data Source:[5]

Figure 2.1 and 2.2 demonstrates different production units and prices in the Nord pool system. Hydroelectric production, the dominant production unit, is highly reliable on weather conditions. Nearly 75% of electric power production relied on hydropower (57%) and nuclear generators (18%), which have lower marginal costs of production. These two traditional production units are used during normal conditions. Gas-turbine and thermo production display are only used for the peak period, for their and high efficiency in generating power and, most importantly, relatively high marginal costs. Therefore, if the increase in demand exceeds a certain amount that traditional production units cannot fulfill, the spot price will rise due to the demand curve shifting to the right [6].

2.2 Data set

In this thesis, we use the data set from the Nord Pool power market. This data set is formed of daily average spot prices calculated from hourly-recorded electricity prices in Eur per Megawatt Hour (Eur / MWh) across eight years, starting from 1st January 1999 to 26th January 2007, comprising 2948 data points.

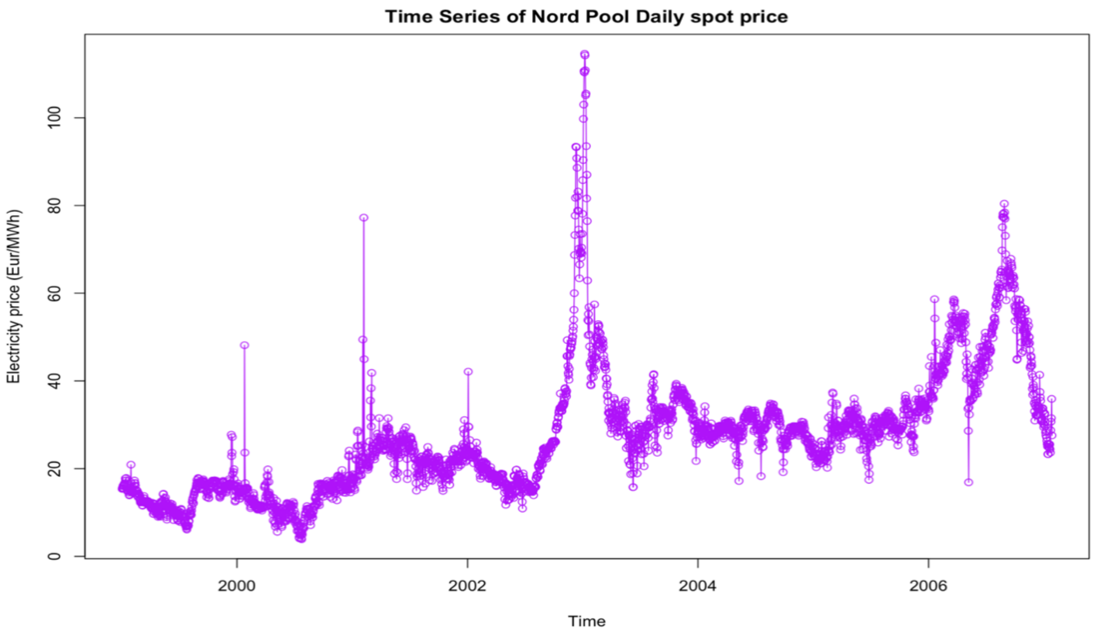


Figure 2.3 Nord Pool Daily Spot Price Time Series Data from 1999 to 2007

From Figure 2.3, irregular volatility and several short-lived price spikes can be observed. This time series is non-stationary and overall shows an increasing trend. The highest price spike was 114.61 Eur/MWh, while the lowest price spike was 3.89 Eur/MWh. Using R command we have, the median price was 26.23 Eur/MWh, and the average price was 27.48 Eur/MWh. Therefore, it is easy to see that prices can upsurge and shrink within a very short period.

2.3 Characteristics of electricity prices

Unlike other tangible and storable commodities like metal or oil, electricity owns non-storable characteristics. In this case, electricity can only be used in its generating region, so transmitting to the demand-exceeding region is constrained. Each region in the Nord Pool is equipped with generators to satisfy that area's demand for electricity. Though under normal conditions, each region's aggregate capacity of generation is able to meet the aggregate demand, the unpredictable weather conditions, non-storable properties, transmitting constraints would trigger inadequacy of supply, then leading to extremely high spot price value (the demand for electricity is very inelastic) until supply problems are settled [6]. From Huisman, we know that Price spikes " occur due to frictions in demand and/or supply to which electricity producers cannot respond flexible enough." [8]. In Figure 2.3, the price spikes are shown as an upsurge following a sudden fall and then reverting to the mean price level within a very short time.

3 Background of Methodology Used

3.1 Returns (First order differencing)

As we mentioned earlier, Figure 2.2.3 presents that the time series for daily spot price is non-stationary and shows an increasing trend. To make our dataset statistically meaningful, we plot the daily logarithmic returns.

A return describes the relative change in a variable compared to its one previous value. The return is a widely used measure in economics and financial time series analysis for detecting price fluctuations, as we can compare all variables since they are time-independent. Here get a stationary time series by doing trend differencing, using the method of return. So that all observed values are independent of time changes and have well-defined means and stable variances.

The concept behind return is the first-order differencing; differencing acts like a high-pass filter, removing low-frequency components and letting high-frequency signals pass by [9].

In this thesis, the equation of simple return is

$$R_t = \frac{X_t - X_{t-1}}{X_{t-1}} = \frac{X_t}{X_{t-1}} - 1 \quad (3.1)$$

where X_t is the daily-average spot price at time t.

Take log on both sides of the equation, we obtain the log return equation as

$$r_t = \log(R_t) = \log\left(\frac{X_t}{X_{t-1}}\right) = \log(X_t) - \log\,(X_{t-1}) \quad (3.2)$$

where $\log\,(X_t)$ is the price taking natural logarithmic at time t.

Logarithmic return is more widely used in analyzing prices because of its three overwhelming advantages: time additive, variance stabilization（shown in Figure 2.5）and "log-normalization."

For time additive, for example, if we want to get the returns form t to $t+n$, from Equation 3.3 we have the log returns of one time span can be calculated directly by adding each individual log returns.

$$r_{t+n,t} = In\left[\frac{X_{t+n}}{X_t}\right] = In[\frac{X_{t+1}}{X_t} \times \frac{X_{t+2}}{X_{t+1}} \times \ldots \times \frac{X_{t+n}}{X_{t+n-1}}]$$

$$= In\left[\frac{X_{t+1}}{X_t}\right] + In\left[\frac{X_{t+2}}{X_{t+1}}\right] + \cdots + In\left[\frac{X_{t+n}}{X_{t+n-1}}\right] \quad (3.3)$$

Moreover, log returns can be more easily understood because log-returns follow the normal distribution. We will use this property for fitting models later in Section 4.

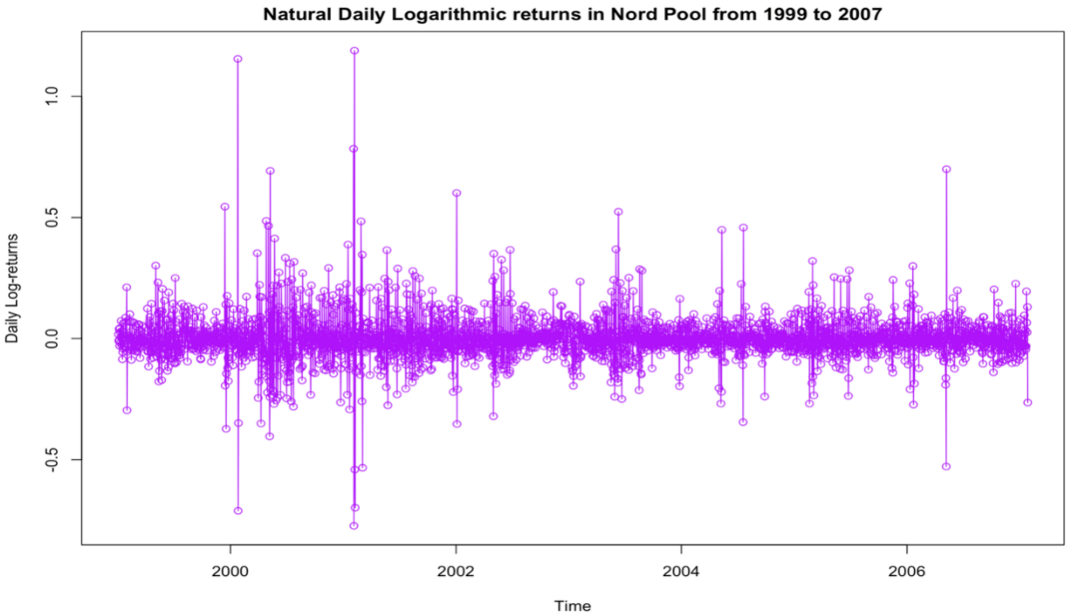


Figure 3.1 Time series of daily logarithmic returns of spot price in Nord Pool from 1999 to 2007

Figure 3.1 is the graph of daily natural logarithmic returns of data observed in the Nord Pool. After transforming, our time series becomes stationary, with fewer variance fluctuations. Both the mean value and the median value approaches 0. Three significant volatilities are observed in 2000, 2001 and 2006 respectively in descending order. The highest value was 1.188913026, and the lowest value was -0.773170223, with a remarkable difference of 1.962083. The most significant price spike in Figure 2.3 normally performs while using daily logarithm returns.

3.2 Power Law

"A Power-law describes a functional relationship between two quantities: one quantity varies as a power of another [10]."

These years, the power-law becomes appealing in extensive scientific research for its and appearance in various natural and man-made phenomena. Its mathematical properties sometimes lead to significant physical consequences, like the well-known Zipf's Law and Pareto's Principal. However, while doing empirical research, instead of being fully interpreted by the power laws at every value of x, usually a studied phenomenon obeys power law for values larger than some minimum only [11](Clauset.A, 2007).

Since the time difference between extreme events in Section 4 is recorded as discrete integer values, we adopt the discrete power-law distribution. The discrete probability distribution function $p(x)$, for a discrete quantity x of integer values, is defined as

$$p(x) = \Pr(X = x) = Cx^{-a} \quad (3.4)$$

where α is a constant parameter called exponent parameter, and C is the normalization constant [11].

Take logarithm on both sides of Equation 2.1, we obtain

$$log\, p(x) = log\, C - \alpha\, log\, x \quad (3.5)$$

Equation 2.2 can be used to perform power-law distribution on the double logarithmic plot, which means that the power-law distribution will act as a straight line of negative slope on log-log axes. So, we can use the method of maximum likelihood and least squares to find the most-fit regression model. We will have a more detailed discussion about this later in Section 4.

3.3 Intermittency

Intermittency is a phenomenon that is widely observed in a dynamic system. Phases altered irregularly from a quiescent(laminar) state into an active(burst) state. In a time series, intermittency can be detected from observations of interchanges between phases with small and large volatilities aperiodically and chaotically.[12]

The term “on-off intermittency" was first raised in 1993; the name of this particular behavior of some chaotic systems derives from the characteristic two-state nature of intermittent signal[13]. The "off" (laminar) state is nearly constant, and it owns the property of long duration. On the other hand, in the "on" (burst) state, the system randomly departs from and returns to the "off" state very quickly[13]. This feature seems to reveal in Figure 3.1, where there is a clear pattern of two states in the time series.

Specific attention is drawn to the distribution of the "off" (laminar) phase. In 1994, Heagy, Platt, and Hammel proved that on-off intermittency holds universal statistical characteristics: parameter-driven. Moreover, the duration of the laminar phases between the bursts exhibits a power-law distribution with the exponent parameter α asymptotes to $-3/2$.

One of the recent experimental studies of on-off intermittency is in semiology, especially on earthquake occurrences. In 2007, Bottiglieri and Godano found that earthquake occurrence exhibited on-off intermittency. The distribution of inactive periods' duration between the class of events follows a power-law distribution with α equals to -3/2[14]. In the conclusion part, they stated that "on-off intermittency behavior is influenced by two variables: the value of the threshold and the length of the bin width Δt" (Bottiglieri & Godano, 2007). Higher threshold values and the bin width Δt will lead to statistical meaningless for detecting patterns, while choosing low values of the threshold value may lead to the observation of chaotic transitions. Figure 3.2 visually presents the findings.

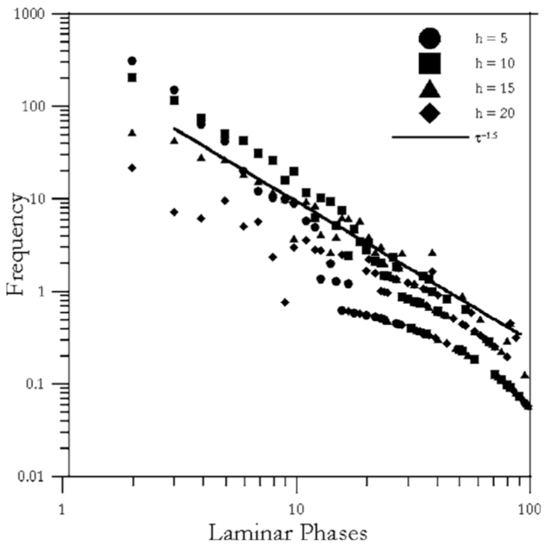


Figure 3.3 Distributions of laminar phases duration of earthquake occurrences for $\Delta t = 1$ day at difference thresholds (Pyragas, 1998) [14]

4 Analysis

This section aims to test whether the points of extreme values follow a discrete power-law distribution with the value of exponent parameter α near to -3/2 and to evaluate the significance of our fitted regression line. First, we plot the neighborhood of the three most significant spikes in season 1 to see how the time series look like in a given period.

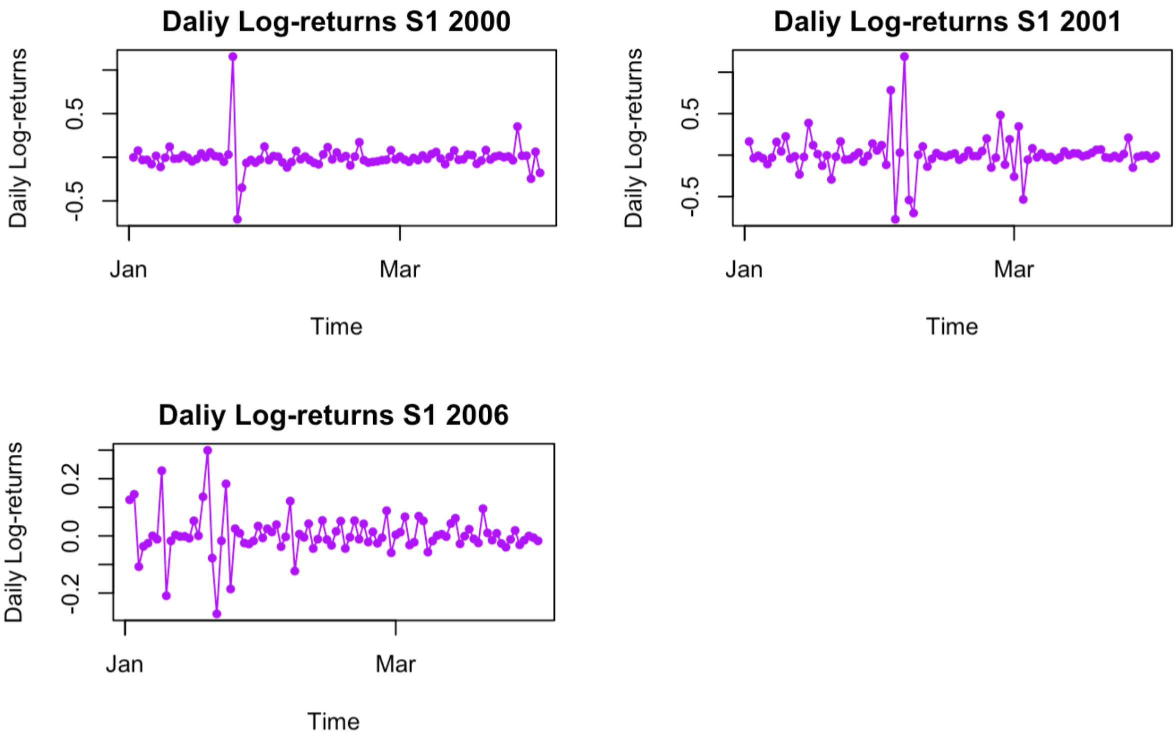


Figure 4.1 Daily log-returns for Season1 in 2000, 2001, and 2006 respectively

Three figures in Figure 3.1 show some possible indications of on-off intermittency in more critical details, especially in 2000 and 2001, turbulent bursts happened and reversed to inactive phases very quickly.

4.1 Time Series Plots

In this thesis, we define the extreme events as the absolute values of adjusted daily log returns which exceed a certain threshold.

To see the extent to which the observed value deviated from the mean value, we adjust our dataset by subtracting the mean of daily log returns from the original data plotted in Figure 3.4.

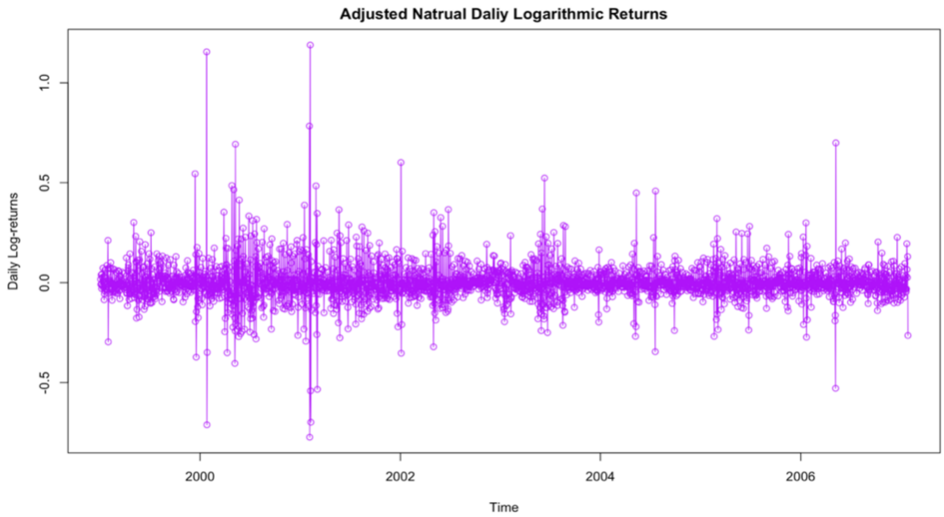


Figure 4.2 Daily natural logarithmic returns with the mean subtracted (deviations)

Figure 4.2 presents the dispersion of the data, which looks very similar to the original time series plot (Figure3.4) since the mean is almost zero.

To make Figure 4.2 more interpretable in visualizing all extreme events, we transfer all negative daily log returns by taking their absolute values and producing the positive plot (Figure 4.3).

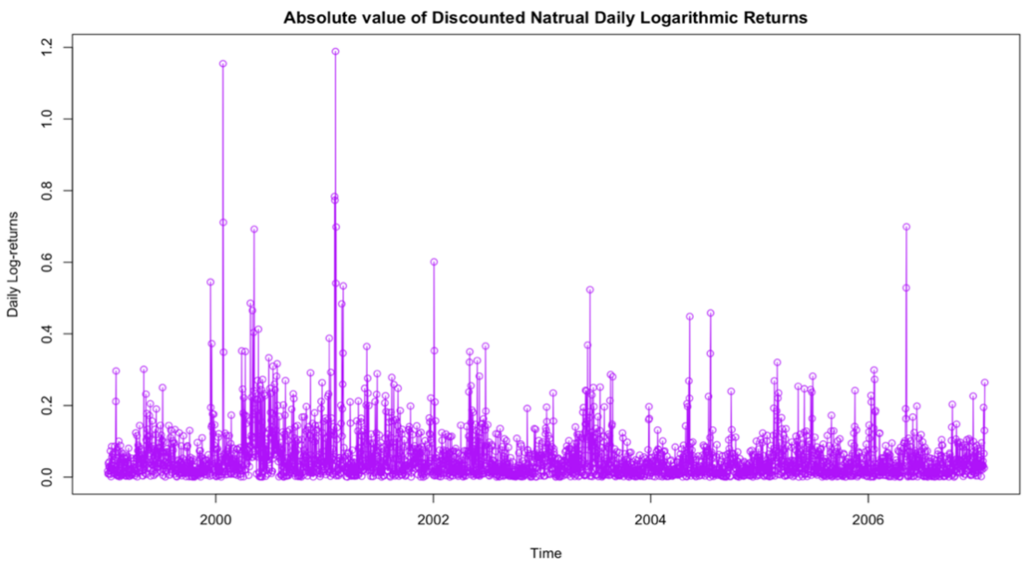


Figure 4.3: Absolute value of adjusted daily natural logarithmic returns

We can choose the threshold value then. It is noteworthy that failure to set an appropriate threshold value would cause problems in the model fitting of statistically insignificant datasets in the following sections. In this case, we choose values above the 75th Quantile. First, we choose 0.15, and the data points for extreme events are presented in Figure 4.4.

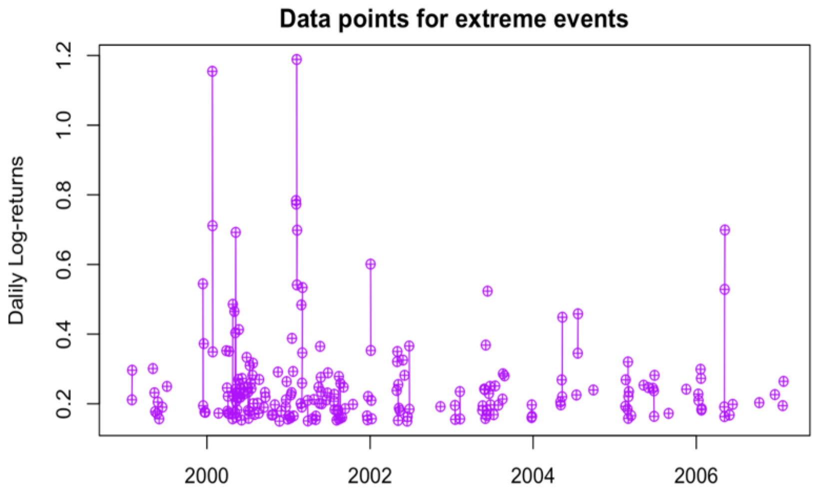


Figure 4.4: Data points that exceed the 0.15 threshold from discounted hourly natural logarithmic returns

4.2 Time Differences analysis

4.2.1 Histogram of Distribution of Time Spans

We aim to determine whether the frequency distribution of the time difference between two subsequent extreme events (upsurge or shrink in daily log-returns) follows a power-law distribution with the exponent parameter α equals or close to -3/2. Thus, the first step is to calculate the time differences between two subsequent bursts. Then we generate a histogram to summarize the frequency of these times spans that occurred. It is noteworthy that while setting the bin widths, some slight differences can be ignored, as they may belong to the same extreme event. In other words, doing so will not trigger the failure to identify what type of distribution our frequency of time differences belongs to.

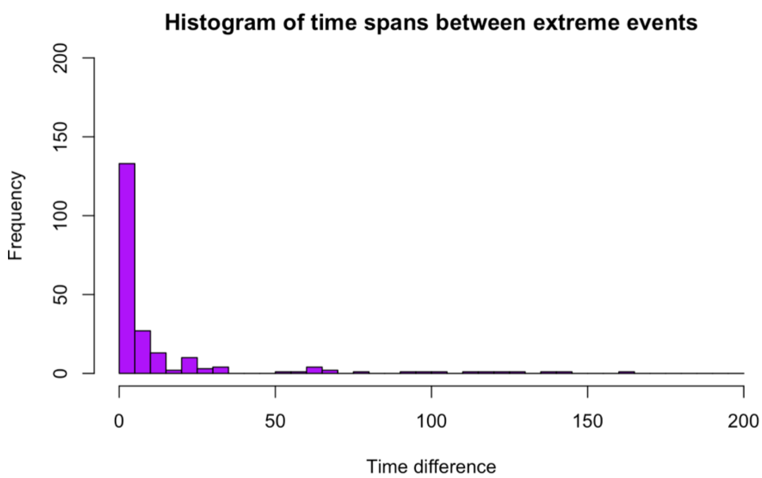


Figure 4.5 Histogram of the time difference of daily log returns above threshold 0.15

In Figure 4.5, we observed that the histogram is heavily positively- skewed: most time differences fall to the first bin, with a few time differences shown in the rest bins.

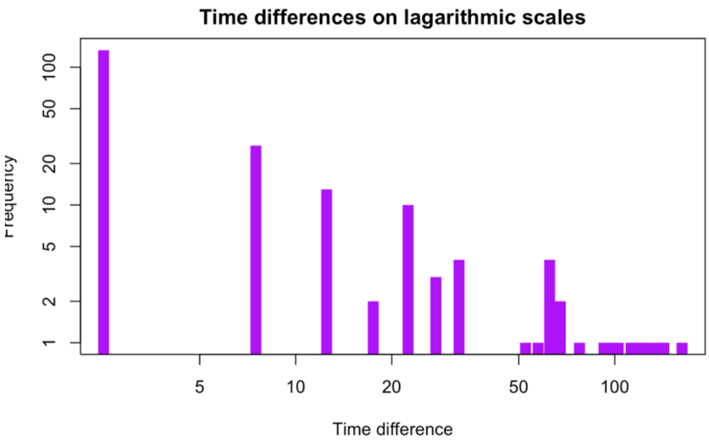


Figure 4.6 Histogram of the time difference of daily log returns above threshold 0.15 on logarithmic scales.

Figure 4.6 presents all the time differences between extreme events on the log-log axes. The frequency decreases as time increases. Figure 4.6 is more visually interpretable than Figure 4.5 because when plotting $log\ p(x)$ against $log\ x$, we will get the result as a straight line with slope e equal to the value of α.

(Note that bins with zero observations are excluded in two figures as $log(0)$ is undefined.)

4.2.2 Simple Linear Regression

In Section 3.2, we mentioned that data follows a power-law distribution if it presents a straight line with a slope equal to α in the double logarithmic plot. Learning from Section 3.3, we know that the on-off intermittency behavior is identified by a power law with $\alpha =$ -3/2. So, to find out whether our spot price data exists an apparent similarity with on-off intermittency, we aim to fit a straight line on the histogram in Figure 4.5 and estimate the scaling parameter α using the gradient of that line.

Learning from equation 3.6, we have the power-law distributions presented as $log\ p(x) = log\ C - \alpha\ log\ x$. In practice, we can fit a linear regression line for the data spots on the histogram in Figure 4.7, and the intersection point of this line (if we extend it) on the $\log(y)$ axis is $\log(C)$, where C is the normalization constant. Moreover, the slope of this line is seen as the estimate of parameter α. Here we only have one explanatory variable, so we use a single linear regression model $y = \beta_0 + \beta_1 x$. The natural log value of time difference is the in independent variable x, and the counts being the dependent variable y.

In constructing a straight line of which each value y on the straight line is as close as possible to the corresponding actual observed data y_i for each x_i, we apply the least-squared method. This method chooses the line where the sum of the squared residuals is minimized. The residual is defined as $\varepsilon_i = y_i - \beta_0 - \beta_1 x_i$, so we are aiming to minimize

$$\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Here we are using the R command lm( ) to create the simple regression line, also called least-squared line $y = \widehat{\beta_0} + \widehat{\beta_1} x$, for $\widehat{\beta_0}$ *and* $\widehat{\beta_1}$ represents are estimated parameter.

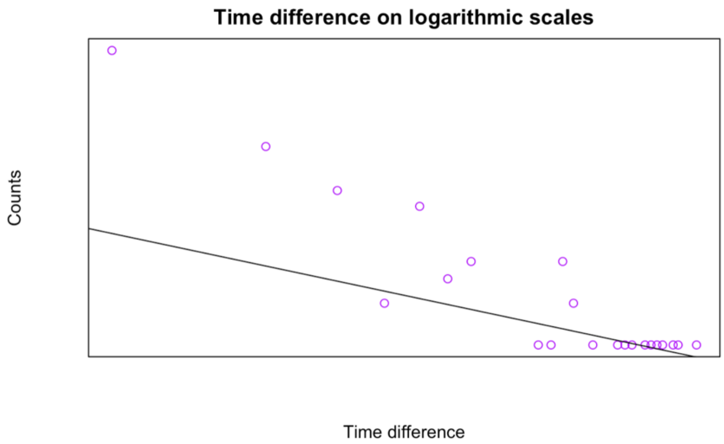


Figure 4.7 Histogram with a fitted regression line

Using R command, we produce the scatter plot of time differences on log-logs scales with the dotted line illustrates the prediction of counts in each time difference (Figure 3.7). The prediction model gives $\hat{y} = 2.304 - 1.134 x_t$. The slope of the line is -1.134, not very close to -3/2, but we can use R command summary () to decide whether the data owns on-off intermittency characteristics.

4.2.3 Evaluating the model fitting

This section will analyze the fitted model's linearity, normality, and independence from the thorough report produced by the R command summary( ).

```
Call:
lm(formula = lgy ~ log10(h$mids))

Residuals:
     Min       1Q   Median       3Q      Max
-0.59366 -0.06844  0.04822  0.12007  0.33411

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.3039     0.1821   12.65 5.31e-11 ***
log10(h$mids)  -1.1337     0.1026  -11.05 5.75e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2229 on 20 degrees of freedom
  (11 observations deleted due to missingness)
Multiple R-squared:  0.8593,    Adjusted R-squared:  0.8522
F-statistic: 122.1 on 1 and 20 DF,  p-value: 5.754e-10
```

Figure 4.8 summary( ) output in R corresponding to the linear regression model

Residuals

Recall that in a simple linear regression model, the residual, ε_i, is the difference between the observed values y_i and fitted values $\widehat{y_i}$, presented as $\varepsilon_i = y_i - \widehat{y_i}$. We would like our residuals to perform a symmetrical distribution so that the sum of all residuals is near zero, suggesting the least squared line yields where the differences between predicted and observed values are small enough. Looking at the "Residuals: "part in roughly Figure 4.9, the median is greater than 0, and the absolute value of the minimum residual is greater than that of the maximum residual. We may derive the result that our residual distribution is slightly negatively skewed, not strictly symmetrical. For precision, we visualize this with a normal probability plot.

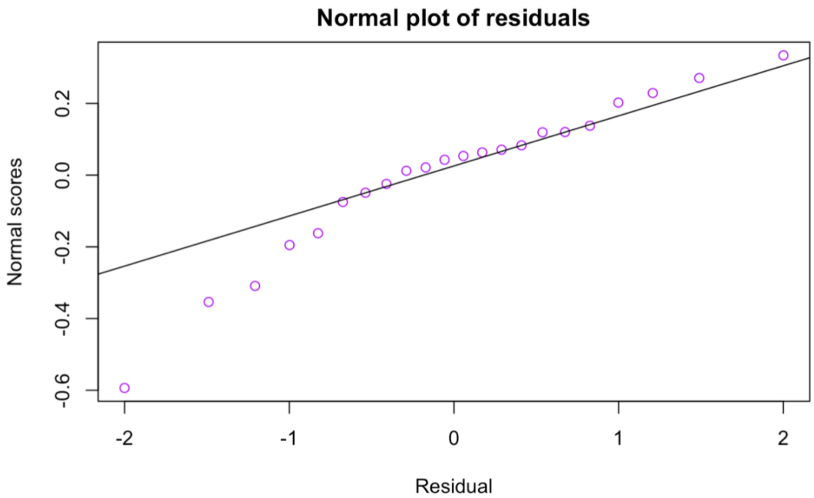


Figure 4.9 The normal Q-Q plot of the residuals (Threshold=0.15)

Nearly all the residuals fall very close to the straight line in the normal probability plot of the residuals (Figure 4.9), except for the five residuals on the left. Therefore, we can say that most residuals follow a normal distribution.

Coefficients

Recall that our estimated model is:

Log density = *intercept* + *slope* × log time difference,

where *intercept* and *slope* will be filling in with the estimated coefficients of the linear regression model. By filling in the estimated coefficients, we have: Log density = *2.3039 – 1.1337* × log time difference. The standard error of each coefficient is an estimate of the coefficient's standard deviation. We can use it to calculate the confidence interval to show the variability in actual slope and intercept.

The t-statistic is calculated by the coefficient divided by the standard value. A large t-value is sought since it indicates a small standard error in comparison to the coefficient. The t-value is used to find the p-value because the p-value is calculated using the t-statistic from the T distribution. P-value tells us how statistically significant our coefficient is to the model. Typically, a p-value equals to or below 0.05 is regarded as significant in practice. In figure 4.9, the p-value for both intercept and the slope is less than 0.001(three asterisks), supporting the extreme significance of this model.

Multiple R-squared

Multiple R-squared (R^2) tells the percentage of variation within the dependent variable (counts) that the predictor variable (log time difference) is explaining. The value of R^2 ranges from 0 to 1; the higher the value, the better the estimated model fits the actual data. For our data, in Figure 4.9, log-time differences illustrate 85.92%

of the variation within the counts, which strongly indicates that our model fits the data very well. Note that as our data are bivariate (we only have one predictor in our model), not multivariate, we do not have to look at the adjusted R-squared.

F-statistic

A hypothesis test is being run on the whole model to test the existence of a correlation between the independent variable and the dependent variable. Mathematically, we set a hypothesis: $H_0 = \text{coefficients for all variables are zero}$; H_1= there is one of them, not zero. A large F-statistic associated with a p-value less than 0.05 generally indicates that we should reject the null hypothesis. In Figure 4.9, the F-statistic is 122.1 with a p-value less than 0.001, which suggests overwhelming evidence against the null hypothesis. In other words, there is no relationship between the counts and the log time difference.

From the results above, we can conclude that the fitted model is *significant* for a threshold value of 0.15. However, since the estimated slope -1.134 is not close enough to -3/2, and dispersions from the normal probability line are seen in Figure 4.9, we continue to find a more convincing threshold value.

4.3 Remodeling by using Different Threshold Value

The first goal of this task is to find the best possible threshold value for identifying on-off intermittency behavior. Then we choose 0.6 as a higher threshold value to see what would happen while fitting the corresponding linear regression models.

We apply the same methods as previous but changing the threshold value to 0.08 and plot the histogram with the least squared line in Figure 4.11. The gradient is now -1.529, extremely close to -1.5. Compared to Figure 4.7, the straight line in Figure 4.11 seems to perform better in fitting the observed data. So, to test our hypothesis, we apply the R command summary() and then compared it to the fitted model using the threshold of 0.15.

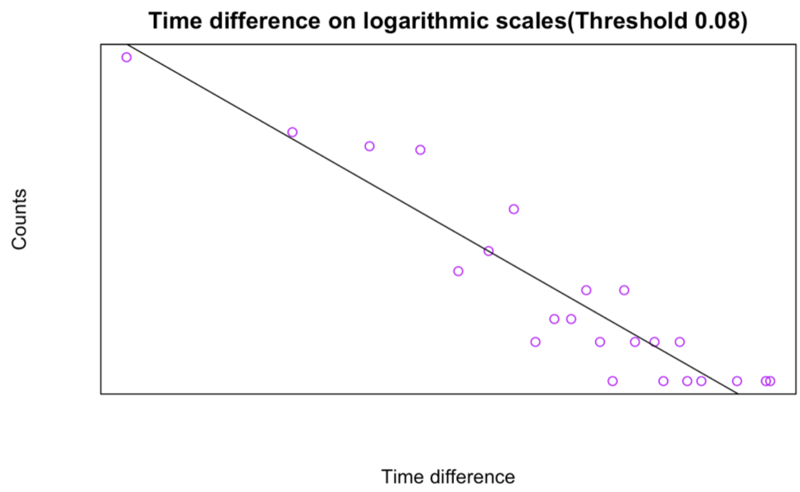


Figure 4.10 Histogram with a fitted regression line (Threshold 0.08)

```
Call:
lm(formula = lgy2 ~ log10(h2$mids))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.49181 -0.18332  0.00116  0.18177  0.47911 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)      2.5915     0.1693   15.31 7.27e-13 ***
log10(h2$mids)  -1.5293     0.1262  -12.12 6.08e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2655 on 21 degrees of freedom
  (13 observations deleted due to missingness)
Multiple R-squared:  0.8749,	Adjusted R-squared:  0.8689 
F-statistic: 146.8 on 1 and 21 DF,  p-value: 6.077e-11
```

Figure 4.11 summary( ) output in R corresponding to the linear regression model (Threshold 0.08))

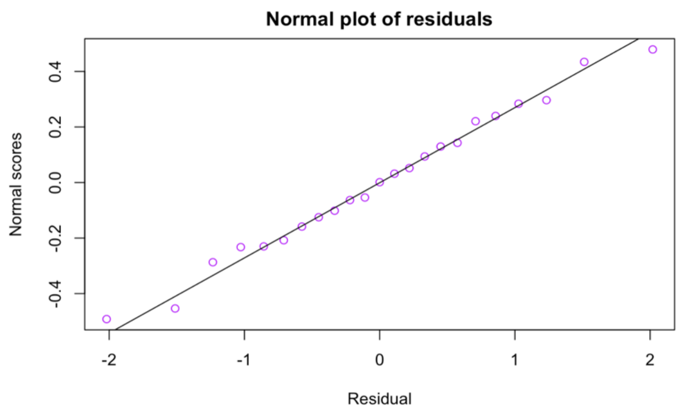


Figure 4.12 The normal Q-Q plot of the residuals (Threshold 0.08)

From the "residuals:" part in Figure 4.12, we can assume that all residuals are symmetrically distributed. Looking at Figure 4.13, the distribution of log values of the residuals is more linear than that of threshold 0.15 since all residuals are located very close to the straight line. Both our p-values for the coefficients are below 0.001 as well, but with a much lower value. Moreover, the value of Multiple R-squared is 87.49%, slightly larger than 85.92%. Then F-statistics is 146.8, larger than 122.1, together with the p-value less than 0.001. Hence, we conclude that the on-off intermittency behavior is more significant when choosing Threshold 0.08.

This time we use a higher threshold of 0.6 to see what would happen when fitting a least-square line for the corresponding histogram.

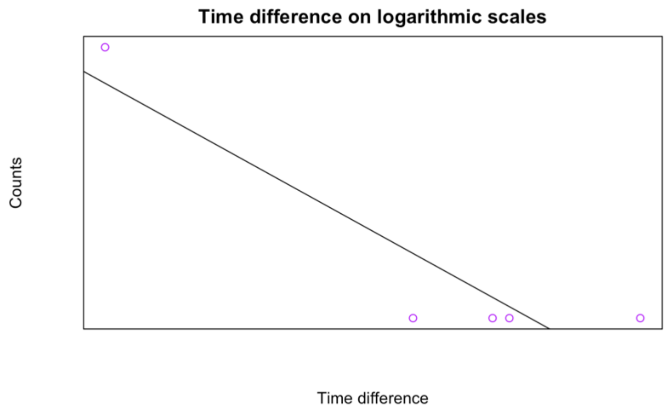


Figure 4.13 Histogram with a fitted regression line (Threshold 0.08)

From Figure 4.11, we can see that only five obverse data exist, as we exclude a large number of returns (Figure 4.4). The number of log time differences is not enough to establish one statistically meaningful regression model if we set a high threshold value.

4.4 Bin Width analysis

In Section 3.3, we mention that apart from the threshold value, the length of bin widths is another variable influencing on-off intermittency behavior on earthquake occurrences. Now, we want to testify whether the electricity price experiences the same characteristics.

In the previous analysis, we fixed the bin widths of 5 to generate histograms shown in Figure 4.7 and Figure 4.10. We want to see how changing the bin widths influences the outcomes quantifiably using argument breaks( ) in Command hist( ). Changing the bin width will affect the grouping of the data, leading to the reshaping of the histogram. Hence the corresponding least-squared line is changed. For example, for threshold 0.15, if we alter the bin width to approximately 20, we will get the slope equals -1.673, which is closer to -1.5 compared to the original slope of -1.134. The new model (Log density = $3.726 - 1.673 \times$ log time difference) is also statistically significant. However, the performance of residual's normality is more unsatisfactory, indicating that increasing the bin width might, to some extent, decrease the variability of the distribution since the same number of data (211 data points) are now packed into a fewer number of bins.

Now we apply a lower threshold value of 0.06. If we set bin width at 5, we get the straight line with gradient -1.803; if we decrease the bin width to 1, then the slope increases to -1.762, bringing closer to -3/2 (Note that 1 is already the minimum bin width). Hence, we can conclude that on-off intermittency can be observed if we choose higher thresholds along with higher values for the bin width or lower thresholds together with lower values of bin width.

5 Conclusions

This project concludes by arguing that the extreme values of electricity daily spot prices in the Nord Pool comply with on-off intermittency characteristics. To derive this conclusion, we used the dataset comprising daily spot price in Nord Pool over eight years, containing 2948 data points. After taking natural logarithmic returns of the dataset, our stationary time series reveals several short-lived spikes alongside a stable state with nearly constant variance, indicates the possible existence of on-off intermittency behavior.

On-off intermittency holds a universal property that the duration of laminar phases between two subsequent bursts follows -3/2 power-law distribution. In this case, we analyzed the time difference between extreme events by varying different values of threshold and lengths of bin width, then plotting the corresponding histogram about the frequency distribution of time differences with fitted regression line on log-log axes. We found that our daily spot price exhibits on-off intermittency behavior most statistically significant when choosing threshold value 0.08 with bin width 5. Those analyses lead to the following conclusion: on-off intermittency behavior can be seen in the daily electricity spot price when proper values of threshold and bin with are chosen.

Ideally, our findings can be replicated in future studies, where identifying and forecasting intermittency in spot price market data are sought, even though our statistical model's simplicity may trigger some limitations.

Appendix

This section presents the R code used to produce output for analysis and generate each graph.

```
```{r}
# Load the data set
setwd('/Users/Karmuaaaaa/Desktop')
data = read.table("BEUR.txt")

start = as.POSIXlt("1999-01-01 00:00:00")
end = as.POSIXlt("2007-01-26 23:00:00")
addtime = difftime(end,start,units = "hours")
addtime
X = ts(data)
htime<- strptime("1999-01-01 00:00:00","%Y-%m-%d %H:%M:%S")+3600*1:70752
```
```

```
```{r}
#https://community.rstudio.com/t/how-to-calculate-daily-average-of-hourly-data-and-for-different-variables-at-the-same-time/42784
install.packages('lubridate')
install.packages('dplyr')
library('lubridate')
library('dplyr')

DF <- data.frame(DateTime = c(seq.POSIXt(start,end,by = "hour")),
                 Price = X)

daytime=difftime(end,start,units="days")
dtime<- strptime("1999-01-01","%Y-%m-%d")+86400*1:2948
DF <- DF %>% mutate(Day= day(DateTime), Month = month(DateTime),Year=year(DateTime))
Daystats <- DF %>% group_by(Year,Month,Day) %>% summarize(DayAvg = mean(V1))
dstats <- Daystats$DayAvg
dlogr <- diff(log(dstats),lag=1)

# Figure 2.2 Nord Pool Daily Spot Price Time Series Data from 1999 to 2007
plot(dtime,dstats,xlab="Time",type="o", ylab="Electricity price (Eur/MWh)", main="Time Series of Nord Pool Daily spot price",col="purple")

# Figure 3.1 Daily natural logarithmic returns of Nord Pool spot price from 1999 to 2007
plot(dtime,c(NA,dlogr),xlab="Time",type="o", ylab="Daily Log-returns", main="Natural Daily Logarithmic returns in Nord Pool from 1999 to 2007",col="purple")

mean(dlogr)
fivenum(dlogr)
max(dlogr)-min(dlogr)
```
```

```{r}
date1 =as.Date("2000-01-01","%Y-%m-%d")
date2= as.Date("2000-03-31","%Y-%m-%d")
df1 = difftime(date2,date1,units="days")
time1 <-as.POSIXlt(date1)+86400*1:91
length(time1)
length(dlogr[366:456])

date7 =as.Date("2001-01-01","%Y-%m-%d")
date8= as.Date("2001-03-31","%Y-%m-%d")
df7 = difftime(date2,date1,units="days")
time7 <-as.POSIXlt(date7)+86400*1:90
length(time)
length(dlogr[751:782])

date5 =as.Date("2006-01-01","%Y-%m-%d")
date6= as.Date("2006-02-01","%Y-%m-%d")
df5 = difftime(date2,date1,units="days")
time5 <-as.POSIXlt(date5)+86400*1:90
length(time)
length(dlogr[732:821])

# Figure 4.1 Daily log-returns for Season1 in 2000, 2001 and 2006 respectively
par(mfrow=c(1,2))
plot(time1,dlogr[366:456],xlab="Time", ylab="Daily Log-returns", main="Daliy Log-returns S1
2000",col="purple",type="o",pch=20)

plot(time7,dlogr[732:821],xlab="Time", ylab="Daily Log-returns", main="Daliy Log-returns S1 2001
",col="purple",type="o",pch=20)

plot(time5,dlogr[2558:2647],xlab="Time", ylab="Daily Log-returns", main="Daliy Log-returns S1
2006",col="purple",type="o",pch=20)
```

```{r}
dmean <-mean(dlogr)
Subtr =dlogr -dmean
Abs = abs(Subtr)

# Figure 4.2 Daily natural logarithmic returns with the mean subtracted (deviations)
plot(dtime, c(NA,Subtr),type="o",xlab="Time",ylab="Daily Log-returns",main="Adjusted Natrual
Daliy Logarithmic Returns", col="purple")

# Figure 4.3: Absolute value of adjusted daily natural logarithmic returns
plot(dtime, c(NA,Abs),type="o",xlab="Time",ylab="Daily Log-returns",main="Absolute value of
Discounted Natrual Daily Logarithmic Returns", col="purple")

```

```{r}
# Figure 4.4: Data points which exceed the 0.15 threshold from discounted hourly natural
logarithmic returns
Eabs<-Abs
Eabs[Abs<0.15] <-NA
plot(dtime,c(NA,Eabs),type="o",xlab="Time",ylab="Dalily Log-returns",main="Data points for
extreme events", col="purple",pch=10)

```

```{r}
df1 <- na.omit(data.frame(Eabs))
df2 <- cbind(DIFF = rownames(df1),df1)
rownames(df1)= 1:nrow(df1)
index <- as.numeric(as.character(df2[,1]))
tail(index,-1)- head(index,-1)
Data<- tail(index,-1)-head(index,-1)

#Figure 4.5 Histogram of the time difference of daily log returns above threshold 0.15
hist(Data,xlim=c(0,200),ylim=c(0,200),breaks=seq(from=0,to=200,by=5), xlab="Time difference",
main ="Histogram of time spans between extreme events",col="purple",)
```

```{r}
h <- hist(Data, plot= F, breaks =40)
h$mids
h$counts

#Figure 4.6 Histogram of the time difference of daily log returns above threshold 0.15 on
logarithmic scales
plot(h$mids, h$counts,log="xy",col="purple",pch=20,xlab="Time
difference",ylab="Frequency",type="h",lwd=10,lend=3, main="Time differences on lagarithmic
scales")
```

```{r}
# Figure 4.7 Histogram with a fitted regression line
plot(log10(h$mids),log10(h$counts),xaxt="n",yaxt="n",xlab="Time difference",
ylab="Counts",main="Time difference on logarithmic scales", col="purple")
lgy<-log10(h$counts)
lgy[mapply(is.infinite,lgy)] <- NA
reg <-lm(lgy~log10(h$mids))
abline(reg)
reg
```

```{r}
# Figure 4.8 summary( ) output in R corresponding to the linear regression model
# Figure 4.9 The normal QQplot of the residuals (Threshold=0.15)
qqnorm(resid(reg),xlab="Residual",ylab="Normal scores",main="Normal plot of residuals",
col="purple")
qqline(resid(reg))
summary(reg)
```

Note that the rest of the analyses are done by simply repeating the R command starting from Figure 4.4.

References

[1]Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*(4), 1030–1081. https://doi.org/10.1016/j.ijforecast.2014.08.008

[2] Heagy, J. F., Platt, N., & Hammel, S. M. (1994). Characterization of on-off intermittency. *Physical Review E*, *49*(2), 1140–1150. https://doi.org/10.1103/physreve.49.1140

[3] Erzgräber, H., Strozzi, F., Zaldívar, J.-M., Touchette, H., Gutiérrez, E., & Arrowsmith, D. K. (2008). Time series analysis and long-range correlations of Nordic spot electricity market data. *Physica A: Statistical Mechanics and Its Applications*, *387*(26), 6567–6574. https://doi.org/10.1016/j.physa.2008.07.030

[4] Oskar Knapik, 2017. "**Modeling and forecasting electricity price jumps in the Nord Pool power market**," CREATES Research Papers 2017-07, Department of Economics and Business Economics, Aarhus University.

[5] *The main arena for trading power*. (n.d.). Retrieved 5th May 2021, from https://www.nordpoolgroup.com/the-power-market/Day-ahead-market/ [accessed 27th, April]

[6] Geman, H., & Roncoroni, A. (2006). Understanding the fine structure of electricity prices*. *The Journal of Business*, *79*(3), 1225–1261. https://doi.org/10.1086/500675

[7] Weron, R., & Zator, M. (2014). Revisiting the relationship between spot and futures prices in the Nord Pool electricity market. *Energy Economics*, *44*, 178–190. https://doi.org/10.1016/j.eneco.2014.03.007

[8] Huisman, R. (2008). The influence of temperature on spike probability in day-ahead power prices. *Energy Economics*, *30*(5), 2697–2704. https://doi.org/10.1016/j.eneco.2008.05.007

[9] William, Y(2021), Lecture Notes: 2.1.6 Trend differencing, Lecture Notes, Time Series, MTH6139/MTH6139P, Queen Mary University of London

[10] Contributors to Wikimedia projects. (2021b, 7th April). *Power law*. Wikipedia. https://en.wikipedia.org/wiki/Power_law [Accessed 4th May]

[11] Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law distributions in empirical data. *SIAM Review*, *51*(4), 661–703. https://doi.org/10.1137/070710111

[12] Platt, N., Spiegel, E. A., & Tresser, C. (1993). On-off intermittency: A mechanism for bursting. *Physical Review Letters*, *70*(3), 279–282. https://doi.org/10.1103/physrevlett.70.279 [Accessed 23rd April]

[13] Pyragas, K. (1998). On-off intermittency at the threshold of weak synchronization of chaos. *Chaos, Solitons & Fractals*, *9*(1–2), 337–342. https://doi.org/10.1016/s0960-0779(97)00071-4

[14] Bottiglieri, M., & Godano, C. (2007). On-off intermittency in earthquake occurrence. *Physical Review E*, *75*(2). https://doi.org/10.1103/physreve.75.026101

[15] Čenys, A., Anagnostopoulos, A. N., & Bleris, G. L. (1997). Distribution of laminar lengths for noisy on-off intermittency. *Physics Letters A*, *224*(6), 346–352. https://doi.org/10.1016/s0375-9601(96)00845-6

[16] DeGroot, M. H., & Schervish, M. J. (2011). *Probability and statistics*. Pearson Higher Ed.

[17] Vásquez, A. R. (2020, 15th January). *Interpreting summary function output for regression model in R*. Unknown. https://www.researchgate.net/publication/338593763_Interpreting_Summary_Function_Output_For_Regression_Model_In_R [Accessed 3rd May]

[18] *Quick guide: Interpreting simple linear model output in R*. (n.d.). Retrieved 6th May 2021, from https://feliperego.github.io/blog/2015/10/23/Interpreting-Model-Output-In-R [Accessed 4th May]