On-off Intermittency in Spot Price Market Data

Yuemin Xu

Student Number:160863705

MTH6138 THIRD YEAR PROJECT, 2018/2019

Supervisor: Dr Wolfram Just

School of Mathematics Sciences

Declaration of original work

This declaration is made on March 23, 2019

I, Yuemin Xu, 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 or by referencing the author together with the year of publication in round brackets, and
2. Using quotation marks "…", and
3. Explicit mentioning of the sources in the last section of this thesis.

Acknowledgment

I would like to give my sincere gratitude to Dr. Wolfram Just, my supervisor, with extraordinary patience and consistent encouragement, gave me great help by providing necessary resources, advice of great value and inspiration of new ideas.

My heartfelt thanks also go to Prof. Alexander Gnedin, my tutor, who took an interest in my education during my time at Queen Mary University of London. He encouraged me to do this project and helped me a lot in the past three years.

Abstract

This paper investigates on-off intermittency in electricity spot market price, we will observe Nord Pool spot market price from 1st January 1999 until 26th January 2006. First of all, we will learn about Nord pool and characteristic of electricity price. In this thesis, the analysis will be based on returns of spot market price, which will use the hourly logarithm returns. We can find high volatility, price spikes and jump clustering from graphs of hourly log-returns made by R programming. High peaks and low peaks indicate the occurrence of extreme events and we will analyze how extreme events affect electricity prices. The objective is to find that electricity spot prices have on-off intermittency behavior. On-off intermittency behavior follows a universal asymptotic 3/2 power-law distribution, we will explore time differences of extreme prices and find it follows this distribution. Power laws appear widely in economics and finance, physics, social science and computer science, we will review some empirical evidence and theories of the existence of power law distribution. To reveal power-law form of distribution, the best way is plotting histogram on logarithmic scales to indicate a straight-line form of power law distribution, so we will explain log-log plot. In this paper, the methodology of producing power law distributions will be shown and some methods to produce time series plots using R will be displayed.

Table of Contents

1.0 Introduction

A scaling behavior called on-off intermittent system has been reported. It is a chaotic signal in coupled dynamical system, as well as this behavior has uncertainty and unpredictability. We can observe the feature from its name, which simply means, the "off" state is nearly constant that can stay on this position for very long periods of time. The "on" state is a burst, which departs from off state quickly and returns quickly to the off state (Heagy, Platt and Hammel, 1994). This phenomenon also appears in electricity price market, we will investigate through practical data.

According to Čenys, "the on-off intermittency phenomenon can be explained in terms of large fluctuations of the local transverse Lyapunov exponents λ [2, 25, 13]" (Čenys et al., 1996). The Lyapunov exponents is a quantitative index to identify characteristic of system dynamics. If a Lyapunov exponent of a system is larger than zero, there are characteristics of dynamical system in this system, and a larger Lyapunov exponent represents a higher degree of chaos. In a finite interval, an "off" state will occur if λ becomes negative. Therefore, statistical properties of the laminar phases corresponding to the "off" state are determined by the distribution P (λ, T) of the local Lyapunov exponents (Čenys et al., 1996). On-off intermittency is associated with invariant manifolds in transverse directions and transverse Lyapunov exponents can be described stability of invariant manifolds.

There are some differences between on-off intermittency and other intermittent system theories. The first difference is bifurcation, on-off intermittency is caused by regulating a parameter through a local bifurcation, in contrast, the other intermittent behavior theories occurs for fixed parameter values, just beyond the bifurcation point (Heagy, Platt and Hammel, 1994). Secondly, the distribution is different in various type of intermittent system. According Heagy, "In on-off intermittency, the distribution of laminar phases is a $-\frac{3}{2}$ power law for short laminar phases, and eventually turns over to an exponential falloff with n as $n \to \infty$" (Heagy, Platt and Hammel, 1994).

In this paper, we will analyze electricity price in Nord Pool from 01/01/1999 to 26/01/2007, that we will focus on extreme price fluctuations. First, an overview of electricity market and Nord Pool will be given. Second, we will learn about returns and power laws. We will investigate statistical properties of the time intervals between extreme events and we will use programming R to obtain results and figures. To get final finding, the specific methodology will be explained.

2.0 Background

2.1 Nord Pool Spot Market

Nord Pool is Europe's leading power market and offers trading, clearing, settlement and associated services in both day-ahead and intraday markets across nine European countries. More than 30 percent electricity consumption in the Nordic region traded by Nord Pool. Nord Pool was established in 1993 and Nord Pool's spot market was organized as a separate company in 2002, the spot market has played an important role from its establishment.

The spot concept depends on bids for purchase and sale of electricity contracts of one-hour duration that cover all 24 hours of the next day. There are three bidding types, namely hourly bids, flexible hourly bids and block bids (Nordpoolgroup.com, 2003). Spot contracts are fulfilled when the electricity is delivered by the supplier or accepted by the buyer. In a spot contract, it is required to complete payment at the same time as the delivery. In energy sector the electricity purchased is paid for one exchange-trading day before it is delivered (Göß, 2016). The spot market provides marginal cost of different time and different place, as well as electricity demand and electricity supply. At Nord Pool Spot, the purchase orders are aggregated to a demand curve. The sale offers are aggregated to a supply curve. We can see aggregated demand curve and supply curve in Figure 1. And the occurrence of electricity price jumps is often loosely simulated by shocks to the electricity demand or by shocks to an inelastic electricity supply. For example, the demands will increase if there are sudden huge changes in temperature and the supplies could decrease caused by production and system breakdowns, increasing demands and decreasing demands both lead to rising electricity price (Hellström, Lundgren and Yu, 2012).

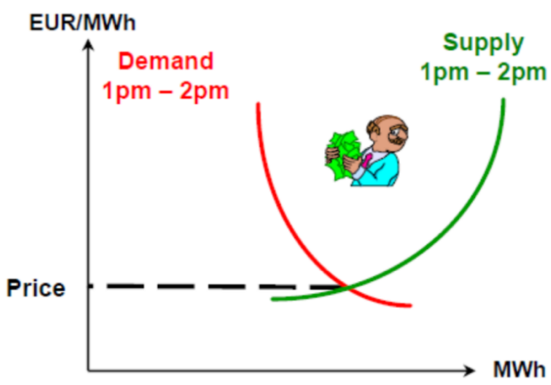

Figure1: The price as a function of MWh (the amount of electricity) for the demand and supply for bids. Source: Nordpoolgroup.com. (2019). [online] Available at: https://www.nordpoolgroup.com/globalassets/download-center/rules-and-regulations/the-nordic-electricity-exchange-and-the-nordic-model-for-a-liberalized-electricity-market.pdf [Accessed 6 Apr. 2019].

2.2 Characteristics of electricity price

Electricity has some special properties and many factors affect electricity prices, causing complicated and abnormal fluctuations. Firstly, electricity is not storable commodity, its subscribed capacity exceeds installed capacity generally. Secondly, electricity should be produced and transmitted at a same time, so it requires a balance of demand and supply for spot electricity trading. Electricity transport must satisfy specific laws to guarantee the grid security. These properties result in a non-predictive electricity price. Therefore, electricity spot price often indicates special characteristics like high volatility, price spikes, seasonality jump clustering.

In this thesis, we will emphasize on electricity spot market data provided by Nord Pool market, starting from the 1st January 1999 to 26th January 2007, which contains 70752 hours. We can see time series of electricity spot price from Figure 2 which has been produced using R programming. The plot shows price fluctuations of each hour, as well as exhibits price spikes, the greatest happened on the 5th February 2001, at 9:00, which was 238.01 EUR/MWh. And the lowest price is 2.3 EUR/MWh, we can conclude that there is huge difference between maximum and minimum. This price curve exhibits very irregular variances and shows chaotic characteristics.

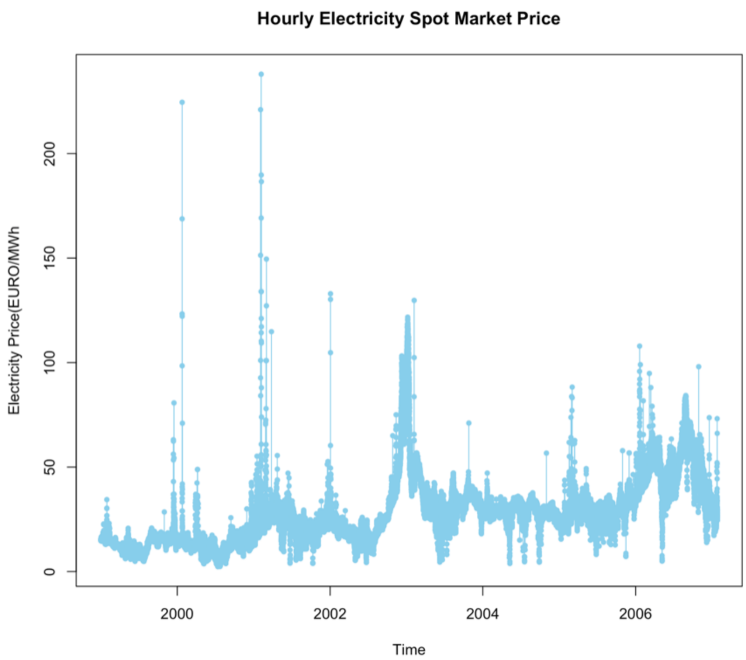

Figure 2: Hourly electricity spot market price is shown as a function of time in the time interval 1999 until 2007 at an hourly sampling rate (that is the distance between two adjacent data points).

2.3 Returns

There are two returns in mathematics, simple returns and logarithm returns respectively. First of all, we learn about two returns and the definitions are shown as follows:

1. Simple return:
   $R_t = (P_t - P_{t-1}) / P_{t-1} = P_t / P_{t-1} - 1$
2. Log return:
   $r_t = \log (P_t / P_{t-1}) = \log (P_t) - \log (P_{t-1})$

where P_t is the price of the asset at time *t*. We define the return from time *t-1* to time *t*. The *log* function here is the natural logarithm. (Portfolioprobe.com, 2010)

Logarithmic return changes will be mainly used to describe changes of electricity price in this thesis. There are some main reasons for using log-return in power market. Firstly, to accommodate characteristics of electricity price dynamics, we need a time-varying conditional variance model, meanwhile, it should contain a price jump component to identify jumps. Second reason is for mathematical convenience, logarithm and exponential are easier to manipulate with calculus contributed to their special product rule. Another factor is time-additive, it said that the two-period log return is equivalent to the sum of each period's log return (profile, 2010). For example, the log return for a year is the sum of the log returns of the days within the year. However, multi-period simple return is the product of the one-period simple returns, which can lead to computational problems for values close to zero (Ageconsearch.umn.edu, 2017).

2.4 Power Law

The study of Heagy, Platt and Hamme found following: at the onset of intermittent behavior, the distribution of laminar phases for a large class of random driving cases exhibits a universal asymptotic $-\frac{3}{2}$ power law (Heagy, Platt and Hammel, 1994). Power law distributions appear in many situations, like physics, computer science, population statistics, social science, finance and so on. In natural world, size of earthquakes volcanic explosion and mountain landslide also conform to power-law distribution. A power-law distribution is a special kind of probability distribution. In general, there are many mathematical methods to express power law, here we display one way to define power laws: $p(x) = cx^{(-\alpha)}$, for $x \geq x_{min}$, where c and α are constants (Clauset, 2011).

In Pareto principle, the economist Pareto discovered that incomes of a small number of people exceed incomes of numerous people. It is also known as the 80-20 rule, which states that for many phenomena, about 80% of the consequences are produced by 20% of the causes (Dunford, Su, Tamang and Wintour, 2014). For example, 20% of institutions own 80% of main fund in stock market, but 80% individual investors only hold 20% fund. In this power law, frequencies of individual incomes which are greater than a certain value has inverse proportion to certain value. The formula is displayed as $P[X \geq x] = x^{(-k)}$, where x is certain value, X is frequency of income and k is constant. The linguist Zipf studied frequency of English words used, which states that, if we rank frequencies of words used from large to small, then the frequencies of certain words are inversely proportional to their rank. The formula can be written as $P(r) = r^{(-\alpha)}$, where r is the rank of word, P(r) is the frequency of word used, and α is constant. And Gutenberg-Richter law discovered the relationship between magnitude and total number of earthquakes.

We now review the empirical evidence of existence of power law, the article that Newman reported mentioned population of city (figure 3). He said populations of city follow power law distributions and provided plots about it.

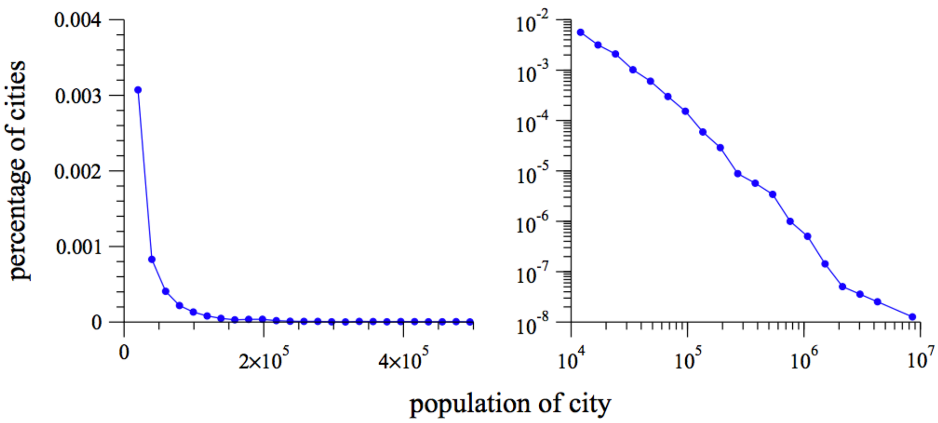

Figure3: population of city. Left: plot of the populations of all US cities with population of 10 000 or more. Right: another plot of the same data but plotted on logarithmic scales.
Source: Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 46(5), pp.323-351.

The right graph indicates the approximate straight-line form of the histogram, which implies that this distribution follows a power law (Newman, 2005). We can find that most US cities have small populations and only a small number of cities have high population, producing a long tail to the right of histogram. We can regard the line in the right plot as a straight line, then we can compute the slope and intercept of this line through making a linear model that provide coefficients (slope and intercept). The slope of this straight line is also exponent α of power law and the intercept is constant c of power law distribution.

The power law is regarded as a useful evidence of self-organized critically (SOC) and a signal that is from a stabilization to chaotic walks, which can be used to forecast phase transition. In reality, there are many interactions between things, when this system reaches to self-organized critically, a small change might cause a huge impact, such as sandpile model. In the sandpile experiment, a pile of sand was built by slow dropping sand, the gradient would be greater and greater following increasing sand. When the gradient of sandpile reached to a critical value, adding sand may cause avalanche. This experiment suggests that the critical state is very sensitive to stimuli for the reason that a small (internal or external) variation can cause a large influence (Hesse and Gross, 2014).

3.0 Analysis

3.1 Time Series Plots

The plot of hourly log-returns based on Nord Pool market data is provided in follows (Figure 4), which exhibits highly volatile across 8 years. We will analyze peaks in the plot and explore what huge impact on the electricity price. The mean value of whole log return data is 6.692666e-06, which is very small and infinite. Using programming R to find the low peak and high peaks, the maximum value of hourly log-return is 1.954708, occurring on 24th January 2000, at 2:00. And the lowest point happened on 7th May 2006, at 2:00, which is equal to -1.263635. The difference between highest value and lowest value is large, it is 3.218343.

Figure 4 indicates three characteristics of natural logarithm returns, including high volatility, price spikes and jump clustering. The occurrence of high peak may be caused by extreme events, such as extreme temperatures, damaging storms, earthquakes. These extreme events happen rarely but they have strong impact. Specifically, the occurrence of price spikes because of some extreme events, when it reaches to a very high temperature, people use air conditioner frequently, the power consumption will be increased in this condition. Using more energy means that natural gas and other powering generation plants will be in greater demand, which will result in electricity price going up. In addition to that, transmission and distribution systems that delivery electricity have maintenance cost, which could be destroyed by severe weather and cost high repairing fees.

We observe that there is on-off intermittency behavior in natural log-returns plot. The values fluctuating around mean value can be regard as the “off” state, these values can remain constant for a long time. And we can consider peak point as the “on” state, which is a burst. We know that occurrence time of highest point and we can find that there are obvious fluctuations over that period of time. To observe change situations during this period clearly, a plot across one week has been produced (figure 5). From 22nd January 2000 at 00:00 to 24th January 2000 at 1:00, natural log-return remained a constant value, it jumped to peak quickly at 2:00 24th January 2000, and then it also dropped sharply in four hours. Huge fluctuations happened during 24th and 25th (log-return increase quickly and decrease quickly), while it returned to a constant value from 26th and keep a long time. Meanwhile, we can see a red arrow between two peak points, this distance can be considered as the duration of a laminar phase.

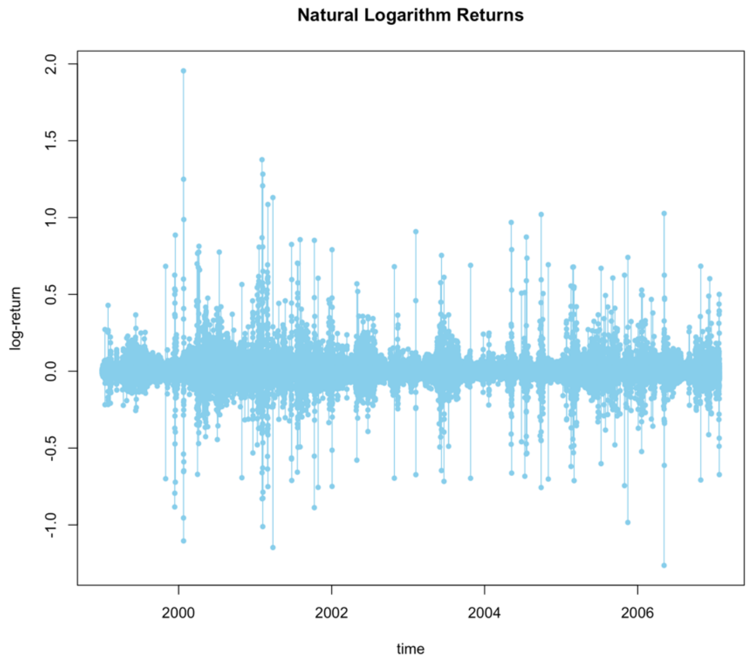

Figure 4: Hourly natural logarithm returns from 1999 to 2007 at an hourly sampling rate (that is the distance between two adjacent data points). Published with R

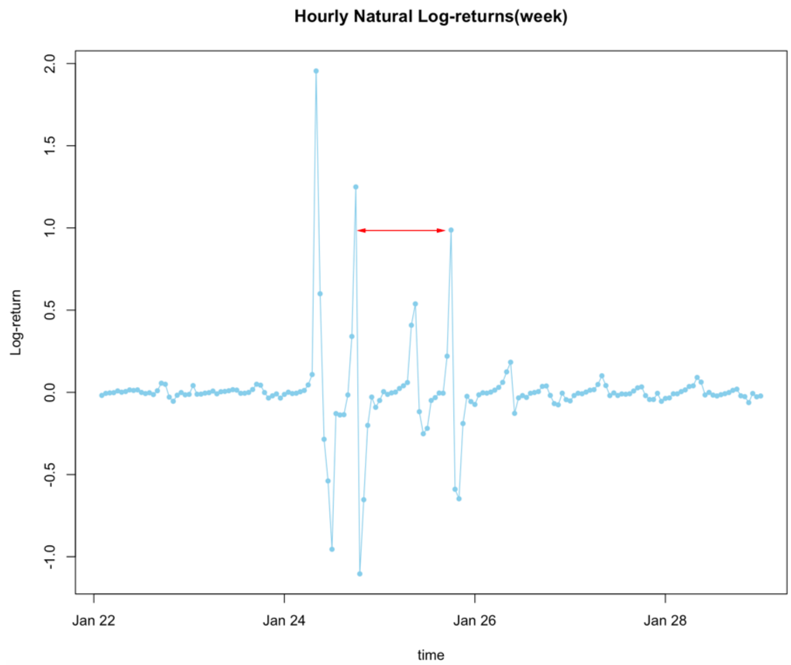

Figure 5: Hourly natural log-returns during a week (from 22nd January 2000 at 00:00 to 28th January 2000 at 24:00). Published with R

3.2 Histogram of Time Difference

In this part, we will emphasize on investigating peak points have on-off intermittency phenomenon, we will analyze occurrence time of peak points and differences of these occurrence time. To produce accurate results for histogram of time difference, we need to do the following steps:

1. Firstly, according to original natural logarithm returns, we use log-return data to subtract its mean value, we will have a new graph subsequently (figure 6). This step makes us know deviations, that is differences between observed value and mean value, indicating dispersion of data.

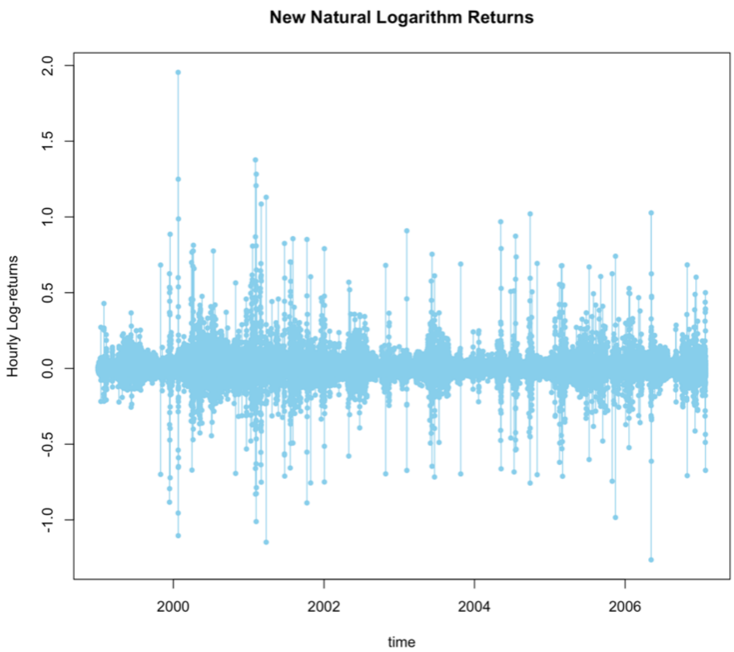

Figure 6: Natural logarithm returns after subtracting the mean value in the time interval 1999 until 2007. Published with R

2. In order to observe time difference of low peak and high peak more clearly, whole data can be changed to absolute value, meaning that all negative data is changed to positive. We look at figure 7.

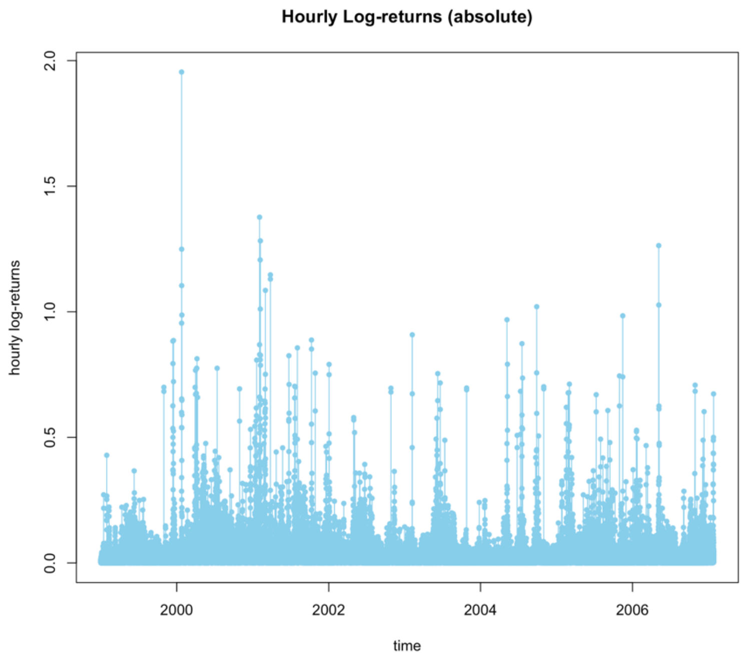

Figure 7: Hourly natural log-returns (absolute) from 1999 to 2007. Published with R

3. Subsequently, we need to pick up those peak points, we first determine a value as standard and choose 0.5 here, then we only pick up the data with 0.5 or more. Now we get a new data set, which only contains 145 points in new data set.

4. The forth step is calculating differences of index (that is time difference).

5. Finally, we produce the histogram of time difference by programming R.

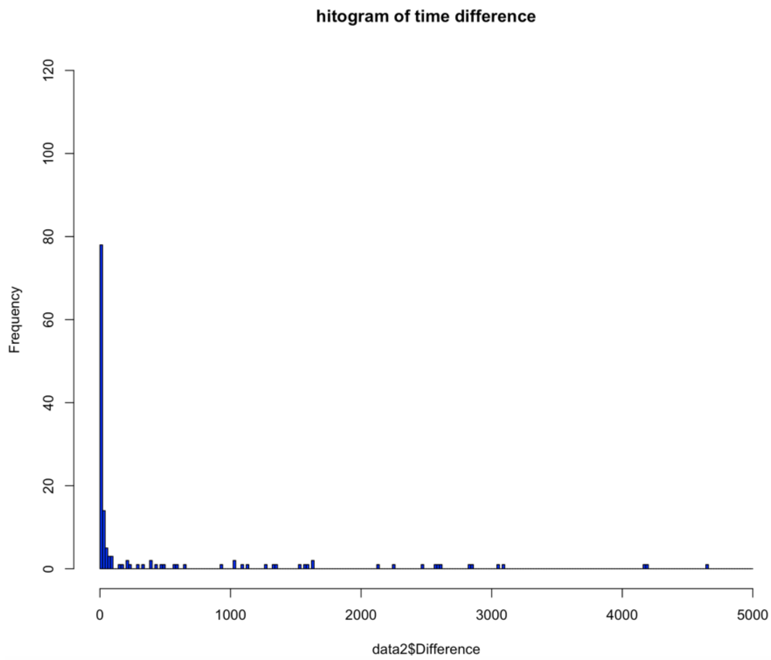

Figure 8 (a): Histogram of time difference of extreme prices (that is the absolute value of natural log-returns larger than 0.5) with break 20. Published with R

From figure 8(a), we can find most differences are in first bin, a small number of differences in other bins. Next, in figure 8(b), we concentrate on the data points that fall on the range of 0 – 400, we can clearly observe this histogram is highly right-skewed. And the red curve is trendline of the histogram, the shape is similar with what we mentioned before, so it corresponds with pawer law.

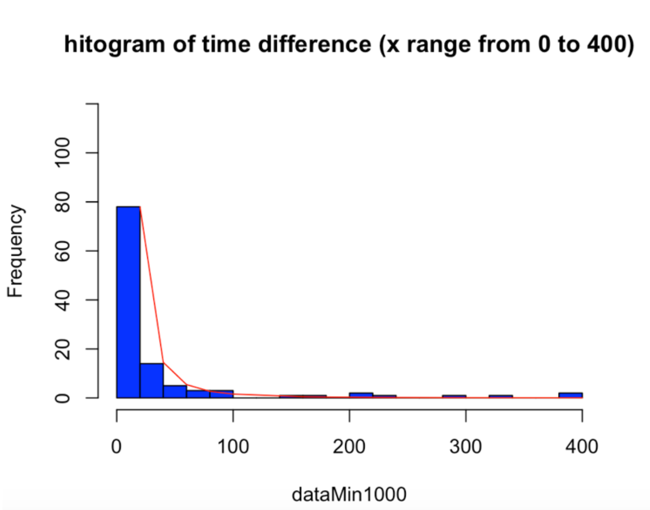

Figure 8(b): Histogram of time difference of extreme prices (that is the absolute value of natural log-returns larger than 0.5) but range of horizontal axis is from 0 to 400 with break 20. Published with R

3.3 Power Law Distribution

When we measure power law distribution, log-log plot will provide a more visualized results for the reason that we can see a straight-line form of the power law distribution in logarithmic scales histogram. As we mentioned before, on-off intermittency behavior follows a universal asymptotic -3/2 power law, we can write its equation as (according to formula mentioned above in power law section):

$$p(x) = cx \wedge (-3/2).$$

Taking the logarithm of both sides, which equals to:

$$\log p(x) = \log (cx \wedge (-3/2)) = \log c - 3/2 \log x.$$

From the equation of this straight line, we can obtain the intercept of line as well as we will know constant *c* of power function.

The histogram of time difference of extreme value is fitted to be a power-law curve, we can see from figure 9 (a), the plot is highly right-skewed, meaning that the smaller time differences have higher frequency of occurrence, producing a long tail towards the right of the plot. Figure 9 (b) indicates log-log plot, it uses same data but plots on logarithm scales. There appears a straight line with formula y = 5.936 -1.584x. The equation of straight line is produced form R programing. Firstly, we need to change horizontal axis and vertical axis to logarithmic scales, then we will fit the data to be a straight line on log-log plot. Secondly, we should produce a linear model of logarithmic data, using lm function in R (actual operation in R will be provided in appendix), then we will obtain the linearity and coefficients (slope and intercept) in this model. We notice the slope of this straight line is -1.584, which is close to -3/2, that is because there are some errors between practical test data and theory, sometimes there exists errors in real tests. Therefore we would say there is on-off intermittency behavior in this model. Meanwhile, the constant *c* of power law can be deduced from the intercept of line, so *c* is equivalent to e^5.936 (that can be equal to 378.4).

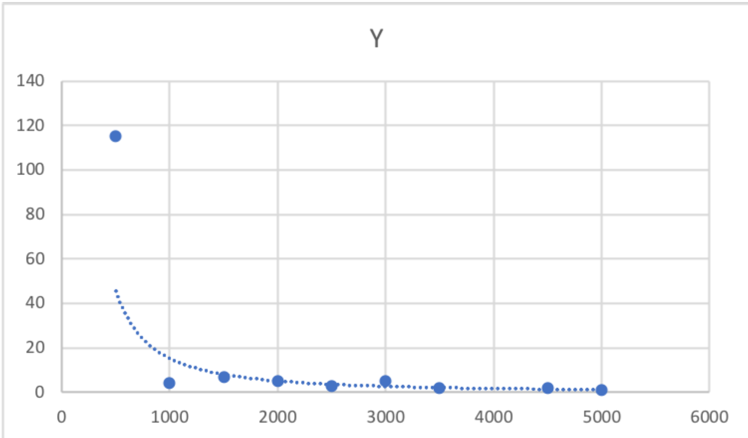

Figure 9(a): histogram of the time difference of all log-returns of 0.5 or more.

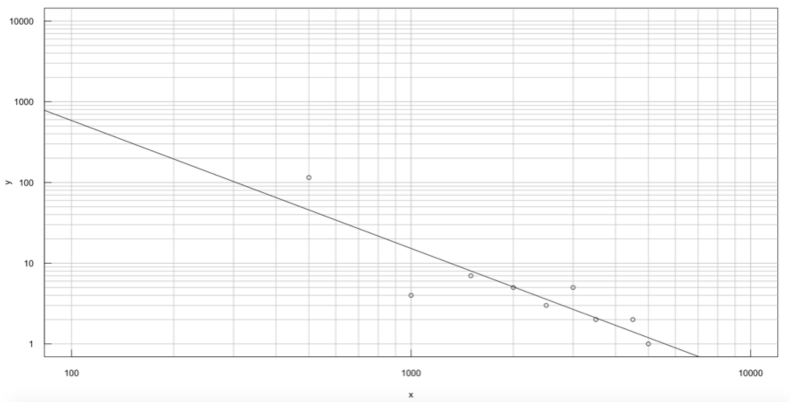

Figure 9(b): histogram of the same data but plotted on logarithmic scales.

We chose 0.5 as the standard value before, and now we can also choose other appropriate number to be the standard value, so we will test 0.6 and 0.2 in next. When we pick up the absolute log-returns which are larger than 0.6, we obtain its linearity with intercept 6.081 and slope -1.574, whose slope is also similar with -1.5. It also indicates that there is on-off intermittency phenomenon in this case. In terms of choosing 0.2 as standard value, we get a linearity with intercept 6.080 and slope -1.771. We observe that there is larger error between exact data and theory. In this situation, there are a lot of data can be regarded as extreme values comparing 0.5 and 0.6. Logically, extreme values happen rarely so that there should be a small number of extreme values. Hence choosing 0.5 or more will make more sense.

4.0 Conclusion

In this thesis, we discovered there exists on-off intermittency behavior in extreme value of electricity spot prices. We used Nord Pool market data which contained hourly electricity price across 8 years from 1999 to 2007. We found that electricity spikes happened rarely and they wer often caused by extreme events, so that extreme events gave huge influence on electricity price. As well as we observed electricity price could increase quickly to the peak points and return to constant state quickly, which indicated the variation tendency of electricity had on-off intermittency phenomenon. In addition to that, we explored that the time difference of extreme price value was said to follow -3/2 power low distribution. Meanwhile, to reveal a better power-law distribution form, we used the logarithmic scales plot to explain the linearity and we obtained the slope of the straight line in log-log plot which was similar to -3/2. Power laws could occur in many fields, we can use logarithmic scales plots to explain relationship between two variates when we are faced with power laws.

However, on-off intermittency is chaotic signal in dynamic system and electricity spot price has chaotic characteristics, so that it is difficult to predict future electricity price. But through this research, we know the changes of electricity prices are abnormal and irregular. Researching electricity spot prices provides useful information for electricity market, which helps traders to formulate a better power policy.

5.0 Appendix

An appendix contains the R command used to generate graphs and outputs which has been indicated in this thesis. Each graph corresponding to a figure that is shown in the main body of this thesis. The following codes are my original work except the codes of Figure 9(a).

Figure 2: Hourly electricity spot market price is shown as a function of time in the time interval 1999 until 2007 at an hourly sampling rate (that is the distance between two adjacent data points).

```
# Read document of data set of spot market price in Nord Pool
setwd("/Users/xukun/Documents")
data <- read.csv("BEUR.csv", header = FALSE)

# Calculating the number of hours
startTime <- as.POSIXlt("1999-01-01 00:00:00")
endTime <- as.POSIXlt("2007-01-27 00:00:00")
time <- difftime(endTime, startTime , units = "hours")

# Generating time series
x <- strptime("1999-01-01 00:00:00", "%Y-%m-%d %H:%M:%S")+3600*1:time

# Plot time series
plot(x, data[, 1], type = "o", pch = 20, xlab = "Time", ylab = "Electricity
     main = "Hourly Electricity Spot Market Price", col = "sky blue")
```

Figure 4: Figure 4: Hourly natural logarithm returns from 1999 to 2007 at an hourly sampling rate (that is the distance between two adjacent data points).

```
# Read the documen data set
setwd("/Users/xukun/Documents")
data <- read.csv("BEUR(1).csv", header = FALSE)

# Calculating the number of total hours
startTime <- as.POSIXlt("1999-01-01 01:00:00")
endTime <- as.POSIXlt("2007-01-27 00:00:00")
time <- difftime(endTime, startTime , units = "hours")

# Generating time series
x <- strptime("1999-01-01 00:00:00", "%Y-%m-%d %H:%M:%S")+3600*1:time

# Calculating the hourly logarithmic returns
z <- data.frame(data)
y <- z[2:70752,]/z[1:70751,]
hourlyReturn <- log(y)

# Plot the hourl log-return
plot(x, hourlyReturn, type = "o", pch = 20, xlab = "time",
     ylab = "Hourly Log-returns", col="sky blue",
     main = "Natural Logarithm Retruns")

# Calculating mean value
result.mean <- mean(hourlyReturn)
print(result.mean)
```

Figure 5: Figure 5: Hourly natural log-returns during a week (from 22nd January 2000 at 00:00 to 28th January 2000 at 24:00).

```
# Define a funciton
plotOneWeek <- function(targetDate)
{
  ###### data of y-axis #####

  # Read document of original data set
  setwd("/Users/xukun/Documents")
  data <- read.csv("BEUR(1).csv", header = FALSE)
  # x1 is defined to be the starting time
  x1 <- as.Date("1999-01-01")
  # x2 is defined to be the target time
  x2 <- as.Date(targetDate)
  # Calculating the time differences between starting time and target time
  deltaHours <- difftime(x2, x1, units = "hours")
  # print index
  print(deltaHours)
  # Define y-axis, and we know that 168 hours per weeek
  y1 <- data.frame(data)
  y2 <- y1[(deltaHours+2):(deltaHours+168),]/y1[(deltaHours+1):(deltaHours+167),]
  yData <- log(y2)
  print(yData)

  ####### data of x-axis ######
  # Calculating the number of hours
  startTime <- as.POSIXlt(targetDate)
  xData <- startTime+3600*2:168
  print(xData)
  # make a graph
  plot(xData, yData, type = "o", xlab = "time", ylab = "Log-return",
       main = "Hourly Natural Log-returns(week)", col = "sky blue", pch = 20)

}

# choose 22rd of January,2000 this day
plotOneWeek("2000-01-22")
```

Figure 6: Natural logarithm returns after subtracting the mean value in the time interval 1999 until 2007. (this command continues the command of figure 4)

```
# Hourly log-returns subtract the its mean value
Zn <- hourlyReturn-result.mean
# Plot
plot(x, Zn, type = "o", xlab = "time", ylab = "Hourly Log-returns", pch =20,
     main = "New Natural Logarithm Returns", col = "sky blue")
```

Figure 7: Hourly natural log-returns (absolute) from 1999 to 2007. Published with R (this command continues the above command)

```
# the negative value is changed to positive
Xn <- abs(Zn)
# plot
mean(x, ...) , type = "o", xlab = "time", ylab = "Xn",
     pch = 20, col = "sky blue")
```

Figure 8(a): Histogram of time difference of extreme prices (that is the absolute value of natural log-returns larger than 0.5), which continue the above command.

```r
# Define g as 0.5, choose all values that are larger than g,
# and get a new data set
g <- 0.5
a <- Xn>g
d <- data.frame(Xn)
d1 <- subset(d, d>g)
setwd("/Users/xukun/study")
write.table(d1,"sample2.csv",sep=",")

# plot the histogram of time differences
setwd("/Users/xukun/study/project")
data2 <- read.csv("hitogram of difference 2.csv", header = TRUE)
h <- hist(data2$Difference, xlim = c(0, 5000), ylim = c(0, 120),
          breaks = seq(from = 0, to = 5000, by = 20),
          main = "hitogram of time difference", col = "blue")
```

Figure 8(b): Histogram of time difference of extreme prices (that is the absolute value of natural log-returns larger than 0.5) but range of horizontal axis is from 0 to 400 with break 20. Published with R

```r
#Plotting histogram
setwd("/Users/xukun/study/project")

data2 <- read.csv("hitogram of difference 2.csv", header = TRUE)
dataOfdiff <- data2$Difference
print(length(dataOfdiff))
dataMin1000 <- subset(data2$Difference, data2$Difference<400)
print(length(dataMin1000))
h <- hist(dataMin1000, xlim = c(0, 400), ylim = c(0, 120),
          breaks = seq(from = 0, to = 400, by = 20),
          main = "hitogram of time difference (x range from 0 to 400)", col = "blue")

#Getting break
x <- h$breaks
length(x)

#Deleting the first break beacause it is zero
x1 <- x[-1]
print(length(x1))

#Getting counts
y <- h$counts
print(length(y))

#Making x1 and y be a data frame
xy <- data.frame(x1, y)

#Obtain the subset of larger than zero
yMax0Data <- subset(y, y>0)

#Fistly, get the index that y is zero
xMin0OIndexData <- which(y==0)

#Secondly, delete the data in x1 that corresponding xMin0OIndexData
x2 <- x1[-xMin0OIndexData]

#Fitting model
fit <- nls(yMax0Data ~ a*x2^b, start = list(a=2, b=1.5))
```

Figure 9(b): histogram of the same data but plotted on logarithmic scales. The log-log plot using a function is based on Petr Keil's Github repository,
source: https://raw.githubusercontent.com/petrkeil/Blog/master/2016_07_05_Log_scales/loglogplot.r

```r
# Read the documen data set
setwd("/Users/xukun/Documents")
data <- read.csv("BEUR(1).csv", header = FALSE)

# Calculating the number of total hours
startTime <- as.POSIXlt("1999-01-01 01:00:00")
endTime <- as.POSIXlt("2007-01-27 00:00:00")
time <- difftime(endTime, startTime , units = "hours")

# Generating time series
x <- strptime("1999-01-01 00:00:00", "%Y-%m-%d %H:%M:%S")+3600*1:time

# Calculating the hourly logarithmic returns
z <- data.frame(data)
y <- z[2:70752,]/z[1:70751,]
hourlyReturn <- log(y)

# Plot the hourl log-return
plot(x, hourlyReturn, type = "o", pch = 20, xlab = "time",
     ylab = "Hourly Log-returns", col="sky blue",
     main = "Natural Logarithm Retruns")

# Calculating mean value
result.mean <- mean(hourlyReturn)
print(result.mean)
```

```r

#Define xData as the break of histogram (using 500 break)
#Define yData as the counts of histogram
xData <- c(500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000)
yData <- c(115, 4, 7, 5, 3, 5, 2, 0, 2, 1)

#Taking log values
x_log <- log10(xData)
y_log <- log10(yData)

#Using log-log function and abline function to plot log-log plot
loglog.plot(xlab="x", ylab="y", ylim=c(1, 10000))
points(x_log, y_log)
df <- data.frame(x_log, y_log)
df1 <- df[-8,]
lm.df <- lm(df1$y_log~df1$x_log)
abline(lm.df)
```

6.0 Reference

1. Heagy, J., Platt, N. and Hammel, S. (1994). Characterization of on-off intermittency. *Physical Review E*, 49(2), pp.1140-1150.
2. Čenys, A., Namajūnas, A., Tamaševičius, A. and Schneider, T. (1996). On-off intermittency in chaotic synchronization experiment. *Physics Letters A*, 213(5-6), pp.259-264.
3. Čenys, A., Namajūnas, A., Tamaševičius, A. and Schneider, T. (1996). On-off intermittency in chaotic synchronization experiment. *Physics Letters A*, 213(5-6), pp.259-264.
4. Heagy, J., Platt, N. and Hammel, S. (1994). Characterization of on-off intermittency. *Physical Review E*, 49(2), pp.1140-1150.
5. Heagy, J., Platt, N. and Hammel, S. (1994). Characterization of on-off intermittency. *Physical Review E*, 49(2), pp.1140-1150.
6. Nordpoolgroup.com. (2003). [online] Available at: https://www.nordpoolgroup.com/globalassets/download-center/annual-report/annual-review_nord-pool-spot_2003.pdf [Accessed 21 Mar. 2019].
7. Göß, S. (2016). *Tutorial power trading 6: The spot market for electricity - Energy BrainBlog*. [online] Energy BrainBlog. Available at: https://blog.energybrainpool.com/en/tutorial-power-trading-6-the-spot-market-for-electricity/ [Accessed 21 Mar. 2019].
8. Hellström, J., Lundgren, J. and Yu, H. (2012). Why do electricity prices jump? Empirical evidence from the Nordic electricity market. *Energy Economics*, 34(6), pp.1774-1781.
9. Portfolioprobe.com. (2010). *A tale of two returns | Portfolio Probe | Generate random portfolios. Fund management software by Burns Statistics*. [online] Available at: https://www.portfolioprobe.com/2010/10/04/a-tale-of-two-returns/ [Accessed 10 Apr. 2019].
10. profile, V. (2010). *Why we use log returns*. [online] Mktlssns.blogspot.com. Available at: http://mktlssns.blogspot.com/2010/04/why-we-use-log-returns.html [Accessed 21 Mar. 2019].
11. Ageconsearch.umn.edu. (2017). [online] Available at: https://ageconsearch.umn.edu/record/265595/files/16_APSTRACT_2017_01-02.pdf [Accessed 21 Mar. 2019].
12. Heagy, J., Platt, N. and Hammel, S. (1994). Characterization of on-off intermittency. *Physical Review E*, 49(2), pp.1140-1150.
13. Clauset, A. (2011). [online] Tuvalu.santafe.edu. Available at: http://tuvalu.santafe.edu/~aaronc/courses/7000/csci7000-001_2011_L2.pdf [Accessed 22 Mar. 2019].
14. Dunford, Su, Tamang and Wintour, R. (2014). [online] Pdfs.semanticscholar.org. Available at: https://pdfs.semanticscholar.org/ef82/dfe7b0ef7a88727636f5ad680a464e33e345.pdf [Accessed 10 Apr. 2019].

15. Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), pp.323-351.
16. Hesse, J. and Gross, T. (2014). Self-organized criticality as a fundamental property of neural systems. *Frontiers in Systems Neuroscience*, 8.