Financial Computing MSc Dissertation MTH778P, 2016/17

On-off intermittency in spot price market data

Mincai Hu, ID 160828748

Supervisor: Dr. Wolfram Just

A thesis presented for the degree of
Master in Sciences in *Financial Computing*

School of Mathematical Sciences
Queen Mary University of London

Declaration of original work

This declaration is made on September 7, 2017.

Student's Declaration: I, Mincai Hu, 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 thanks to my supervisor Dr. Wolfram Just. He introduced this creative topic to me and gave me very detailed explanation of the analysis process. After I finished, he also provided very practical suggestions for the draft. Besides, I want to express my gratitude to my parents and friends for their physical and spiritual support during the process of writing this thesis.

Abstract

On-off intermittency is a kind of intermittent behavior in physical signals. In this thesis, the Nord Pool electricity is the analytic target. Characteristics of on-off intermittency are explored in the Nord Pool prices log return. The distribution of laminar phases and the asymptotic power law of it is analyzed. Finally, with the adjustments of threshold values and bin width of histograms, and using the least square fitting method to estimate the slope and drift, the asymptotic power law is verified.

Preface

The main content of the thesis is the analysis of laminar phases distribution and estimation of slope and drift of asymptotic power law. Chapter 1 is the introduction to fundamental knowledge, including Nord Pool electricity market, the log return of prices, and the concept of on-off intermittency. Chapter 2 is the whole process of data analysis including data preprocessing, generating the distribution of laminar phases, using the least square fitting method to estimate the parameters of the power law. Chapter 3 is the conclusion.

Queen Mary University of London
06$^{\text{th}}$ September 2017

Contents

1 Introduction **6**

1.1 Nord Pool Market 6

1.2 Log Return Of Prices 8

1.3 On-off intermittency 9

2 Distribution of Laminar Phases **10**

2.1 Data Preprocessing 10

2.2 Collection of Laminar Phases 11

2.3 Histogram of Laminar Phases 15

2.4 The Asymptotic Power Law 17

3 Conclusions **25**

Bibliography **26**

Chapter 1

Introduction

In this chapter, we introduce the basic knowledge about the Nord Pool electricity market and the concept of log return in financial time series data. Then we introduce intermittent behavior from physics.

1.1 Nord Pool Market

Nord Pool market was created at 1993 because of the Norwegian Parliament's deregulated act in 1991. At that time the predecessor of Nord Pool Statnett Marked AS was an independent company. At 1996, the market was renamed to Nord Pool ASA and a Norwegian and Swedish joint exchange was founded. Finland has joined the Nord Pool since 1998 and Denmark became a member of it at 2000. After Nord Pool took the sole ownership of UK market in 2014, now it is the largest electricity market in Europe and over 80% of electrical energy in the Nordic area is traded in the Nord Pool.

Because of the particularity of electricity market, to become a member of the Nord Pool, except the financial factor, the Nord Pool also requires the potential participant must have a balancing agreement with the transmission system operator. In that way, the participants have the ability to deliver or get power from the main grid. There are 380 active members of the Nord Pool market.

Nord Pool now provides two types of trading for electrical power: day-ahead and intraday market. In this paper, the day-ahead prices are analyzed. Participants trade electrical power contracts for physical delivery the next day. So it is called the day-ahead market. The traditional contracts are for one-hour long electrical power. The trading mechanism of the market is auctions with bids. Before the 12:00 CET(central Europe time) deadline, members submit their purchase or sell contracts for next day. After the deadline, the area prices and system prices are calculated by aggregating the submitted orders. Area price is under consideration of congestion. Congestion is the situation when the transmission capacity between areas cannot satisfy the requirement. System price is the unconstrained price which is calculated without congestion consideration. In this project, the system prices are used.

The electricity prices in Nord Pool have two typical characteristics, high volatility and a large number of extreme changes in a short time. There are many researches have been done on Nord Pool prices. Long range correlations of Nord Pool electricity prices were analyzed in terms of Hurst exponent in [2]. In [1], extreme value theory is applied and accurate estimates, as well as forecasts, were produced. [8] analyzed the diffusion entropy and mean-reversion in Nord Pool market. Finally proposed a GARCH model for electricity prices. Also, some research focused on the prices forecasting. [6] used an autoregressive model and demand and wind power as exogenous variables to forecast the electricity prices. In [15], the characteristics of electricity prices were analyzed and then a jump diffusion and regime switching model was fitted. [17] used a series of models to forecast electricity prices and compare the performance.

In this project, we are going to use 70751 Nord Pool hourly system prices from January 1999 to January 2007.

1.2 Log Return Of Prices

Unlike other prices analysis, the prices are not used directly in this project. The log return is the natural logarithm of the gross return. Log return is used in this analysis because of its advantage in multiperiod.

Now assume it is at time t-1, the net asset value is P_{t-1}, at time t the net asset value is P_t, so the one-period simple return R_t is:

$$1 + R_t = \frac{P_t}{P_{t-1}} \tag{1.1}$$

If there are n periods, the multiperiod simple return is:

$$1 + R_t[n] = \frac{P_t}{P_{t-n}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times ... \times \frac{P_{t-n+1}}{P_{t-n}} \tag{1.2}$$

The log return or continuously compounded return is the natural logarithm of the one-period simple return:

$$r_t = Ln(1 + R_t) = Ln\frac{P_t}{P_t - 1} \tag{1.3}$$

With the log return, the multiperiod return is turned to:

$$\begin{aligned} r_t[n] &= Ln(1 + R_t[n]) \\ &= Ln[(1 + R_t)(1 + R_{t-1})...(1 + R_{t-n+1})] \\ &= Ln(1 + R_t) + Ln(1 + R_{t-1}) + ... + Ln(1 + R_{t-n+1}) \\ &= r_t + r_{t-1} + ... + r_{t-n+1} \end{aligned} \tag{1.4}$$

Finally, the multiperiod return is the sum of the one period return in the form of log return. The analytic target of this project is the log return of Nord Pool hourly prices.

1.3 On-off intermittency

Intermittent behavior is a kind of common phenomenon in physics. [10] has introduced different kind of intermittency: Pomeau-Manneville I, II, III intermittency. Crisis-induced intermittency is defined in [3]. The on-off intermittency, named after its two states, was defined in [9].

The on-off intermittency has two states: the off state is nearly constant and remain for a long time, while the on state departs from the off state and return back to it in a short time. The long time interval of the off state is called laminar phase. [5] has shown that the distribution of laminar phases obeys a universal asymptotic -3/2 power law.

In this project, we detect the on-off intermittency in Nord Pool prices and try to verify the asymptotic -3/2 power law.

Chapter 2

Distribution of Laminar Phases

In this chapter, we are going to explore the on-off intermittency in Nord Pool prices log return. We will focus on the distribution of laminar phases especially the asymptotic -3/2 power law of it.

2.1 Data Preprocessing

As preparation, the mean value and standard deviation of the log return series are 6.6927×10^{-6} and 0.0379. In [2], space time separation plot is drawn for Nord Pool prices and the horizontal level lines are a symbol of stationary behavior. Unlike other kinds of financial time series data, the Nord Po ol prices are more likely to be considered as stationary time series. Because of the stationary feature, per-example mean subtraction can be used for data preprocessing method. Then we can plot the processed log return time series as showed in Fig 2.1.

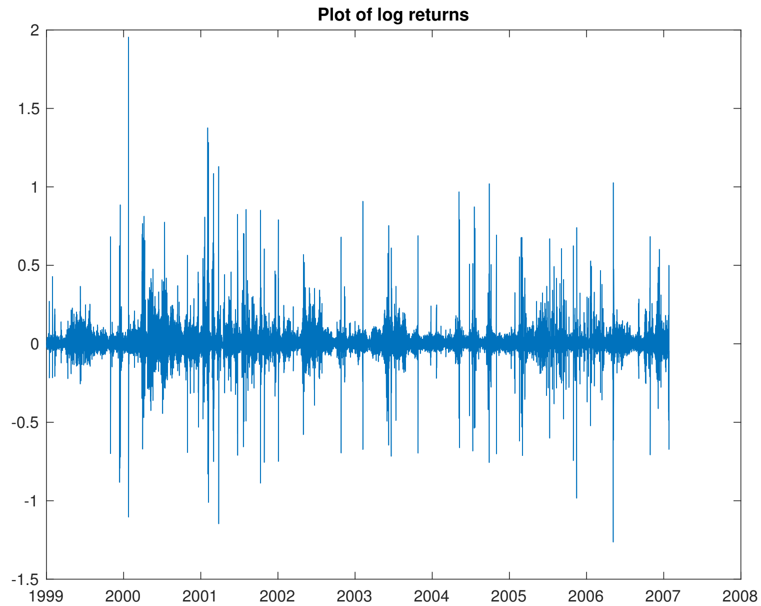


Figure 2.1: Horizontal coordinate shows the time in years while vertical coordinate displays the processed log return of hourly Nord Pool system price from January 1999 to January 2007.

In Fig 2.1, there are lots of bursts in time series and the length of the intervals between them varies.

2.2 Collection of Laminar Phases

As previously mentioned, a laminar phase is the events between adjacent on state or burst in intermittent behavior. To define on and off state of on-off intermittency, a threshold value is needed to set.

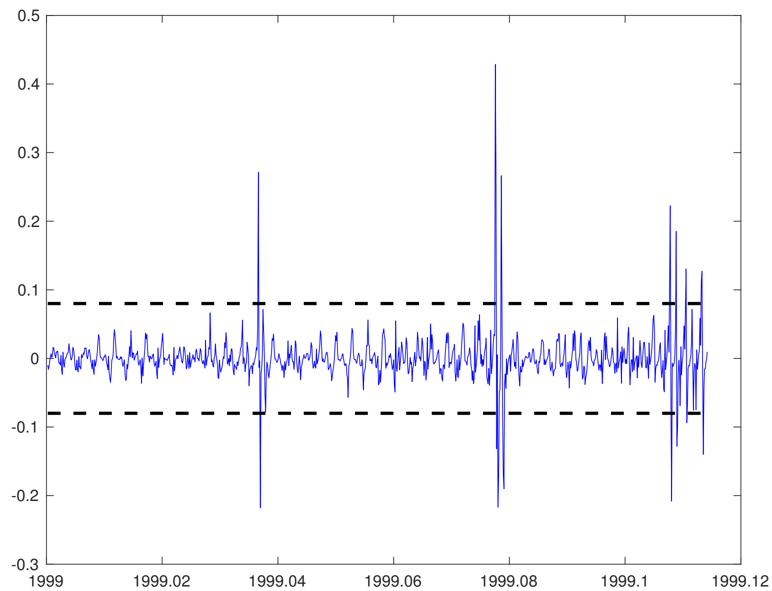


Figure 2.2: Part of samples from Fig 2.1. The threshold value is 0.08.

In fig 2.2, part of the log return has been plotted and the threshold value is set to 0.08 which is the dashed line in Fig 2.2. There are some obvious bursts above the threshold value. They happen instantaneously in extremely short time. They are so-called on states. It is remarkable that after a burst on the positive direction, there is a following burst in the negative direction. The events between on states are laminar phases. Let us take one burst at the positive part as an example. Although the time of the on state is very short, the vertical coordinate already experiences a rising process and a dropping process in a burst.

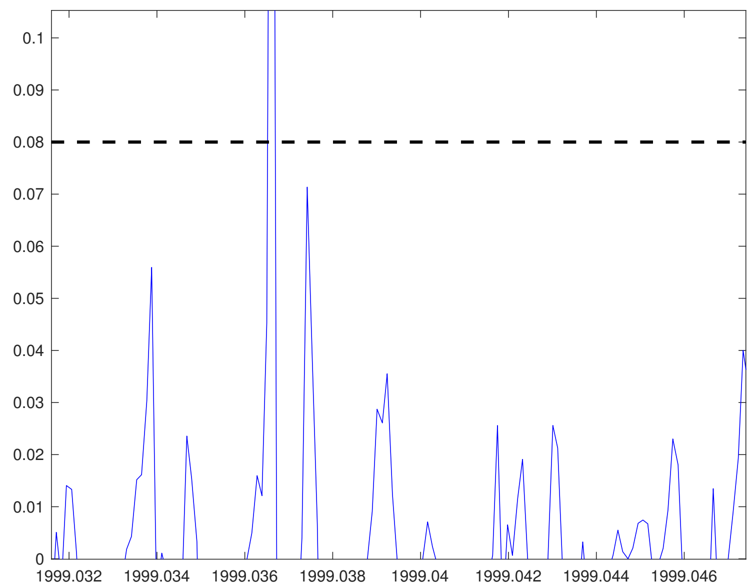


Figure 2.3: Enlarged view of a burst in Fig 2.2.

During these two processes, the plot and the dashed line intersect twice like Fig 2.3 shows. The cross point in the rising process is defined as a rising point while the cross point in the dropping process is dropping point. The absolute value of vertical coordinate of both kinds of points is equal to the threshold value 0.08. Taking the negative part into account, we use the next point's absolute horizontal coordinate to distinguish rising point and dropping point. For example, if a point (x_t, y_t) is a rising point, in time series, $|x_{t+1}| \geq |x_t|$. In the case of a dropping point, $|x_{t+1}| \leq |x_t|$.

With the help of threshold value, the rising points and dropping points are found out. According to the definition of laminar phase, the difference of horizontal coordinate between rising point and drop point of two adjacent bursts is a laminar phase. Imagine there are n bursts in the positive part, the first dropping point is $(x_{d,1}, y_{d,1})$. The rising point of next burst is $(x_{r,2}, y_{r,2})$. The laminar phase is $x_{r,2} - x_{d,1}$.

Actually the above analysis is theoretical. The vertical coordinate of real points in time series are rarely just equal to the threshold value. So we need to calculate the horizontal coordinate of cross points with the assumption their vertical coordinate is equal to the threshold value. The linear equation method is used to solve the cross points.

First we need to find out the points in the neighbourhood of cross points. Every point in the time series is checked. Let us assume the two adjacent points are: (x_{t-1}, y_{t-1}), (x_t, y_t) and the threshold value is: V_t. There are 4 situations which means the criticality happen:

1) y_{t-1} $\leq$V$_t$ & y_t $\geq$V$_t$

2) y_{t-1} $\geq$V$_t$ & y_t $\leq$V$_t$

3) y_{t-1} $\geq$-V$_t$ & y_t $\leq$-V$_t$

4) y_{t-1} $\leq$-V$_t$ & y_t $\geq$-V$_t$

Situation 1) and 2) happen above the horizontal axis while 3) and 4) happen at the negative part. Once these situations happen, the two adjacent points are taken to work out the cross point. It is assumed that the two adjacent points and the cross point are all on the same straight line: $y = ax + b$. With the two adjacent points, the slope of the straight line is: $a = (y_t - y_{t-1})/(x_t - x_{t-1})$. The constant term b is:

$$b = y_t - a \times x_t$$

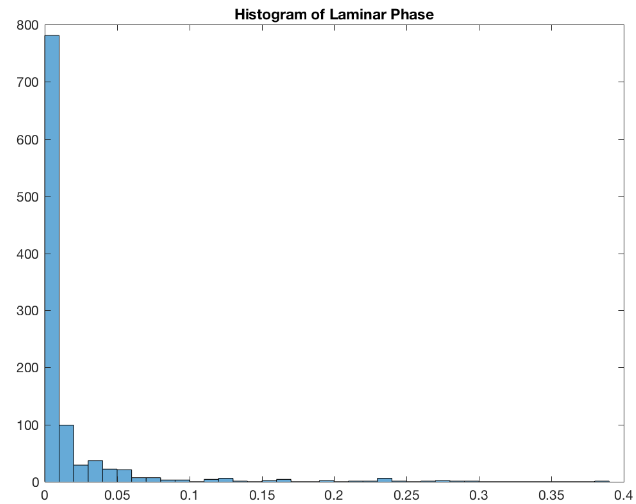
Finally the linear equation $y = ax + b$ has been worked out. Then we use the threshold value to replace the y in the equation and work out the x as horizontal coordinate of cross points. Then we calculate laminar phases according to the definition. Finally, we collect all laminar phases from the Nord Pool log return.

2.3 Histogram of Laminar Phases

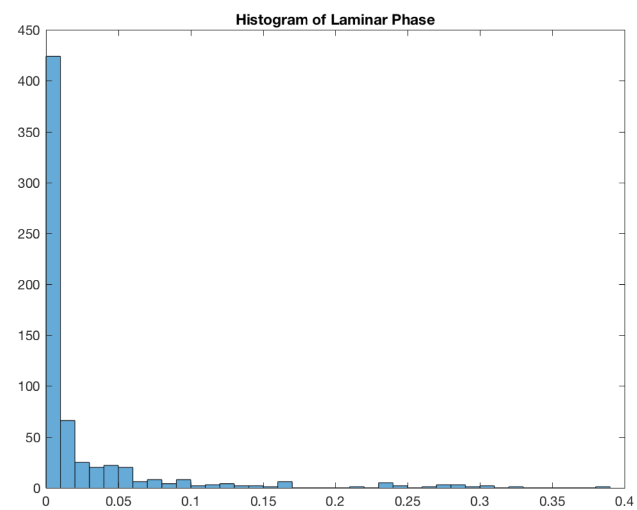
After the data preparation and collection, we can focus on the distribution of the laminar phase.

The histogram is a graphic tool to display the distribution of numerical data. Firstly it divides the range from maximum to minimum to small equal bins. In fig 2.3(a), the range was divided into 39 bins. Each bin is 0.01 wide. Then count the number of samples in each bin. Finally, plot the histogram according to the counted numbers in each bin. We can see that there are most laminar phases in the first bin which range from 0 to 0.01. There are few laminar phases great than 0.15 and the counted number of the following bins are already in the single unit.

Obviously, the threshold value influences the distribution of laminar phase. From Fig 2.3(b) we can know that, when the threshold is 0.20 which is larger than 0.15, the most popular bin is also 0-0.01. But the number is much less than the number in 0-0.01 when the threshold value is 0.15. From the shape of the two histograms, we can say that the distribution of laminar phases is similar to the shape of a power law, which means one variable's change results in a power of another power. The next step of our task is to find out the specific power law.



(a)



(b)

Figure 2.4: Histogram of laminar phases from Nord Pool log return in period from 1999 January to 2007 January. (a) The threshold value is 0.15. (b) The threshold value is 0.20.

2.4 The Asymptotic Power Law

[5] shows that the distribution of laminar phase obeys an asymptotic -3/2 power law in both random and chaotic driving cases. Now we are going to see if this characteristic also exists in Nord Pool log return time series.

Let us take a look at the asymptotic power law first. Suppose there is a constant function P(T), if it satisfies the asymptotic -3/2 power law, it should have:

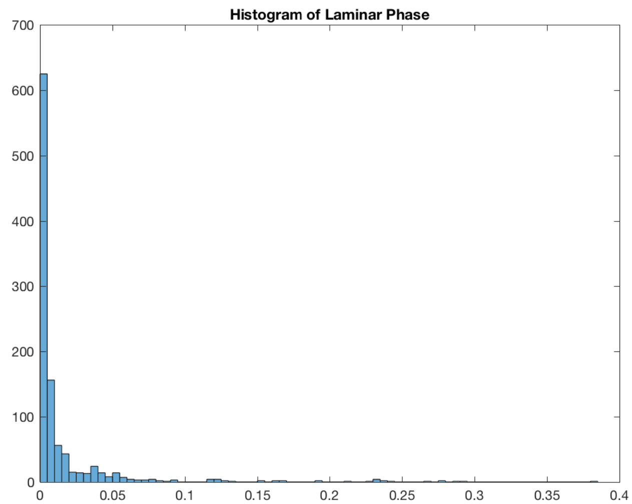
$$P(T) = C \cdot \frac{1}{T^{3/2}} \tag{2.1}$$

Although we know the distribution of laminar phases is approximately a power law, it is still hard to evaluate the unknown C value. We try to take the natural logarithm of both sides of the Eq 2.1:

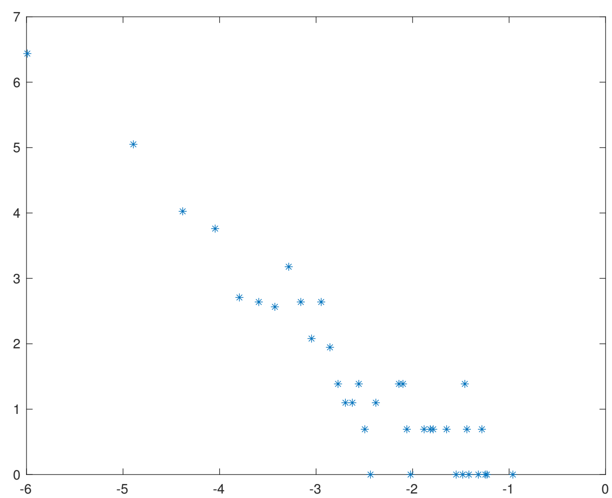
$$\begin{aligned} Ln(P(T)) &= Ln(C \cdot \frac{1}{T^{3/2}}) \\ &= Ln(C) + Ln(\frac{1}{T^{3/2}}) \\ &= Ln(C) - \frac{3}{2}Ln(T) \end{aligned} \tag{2.2}$$

Now the relationship between $Ln(P(T))$ and $Ln(T)$ is a linear relationship. We can estimate the drift $Ln(C)$ and the slope first.

Let us use the case in which the threshold value is 0.15. To ensure the number of bins and enough samples in each bin, the bin width is decreased to 0.005, see the histogram in Fig 2.5(a). In this specific case, the P(T) use number of laminar phases in each bin as samples. Because at each bin, time is a range. For example, there are 625 laminar phases in range 0-0.005. The P(T) is 625 but T is range from 0 to 0.005. In this project, the middle of the range is used as samples of T. So the T is 0.0025 in the first bin. After gathering samples and taking the logarithm of $P(T)$ and T, the data is plotted in Fig 2.5(b).



(a)



(b)

Figure 2.5: (a) The histogram with 0.15 threshold value and 0.005 bin width. (b) The points after double logarithm. The abscissas are $Ln(T)$ while the ordinate is $Ln(P(T))$.

From Fig 2.5(b), we can see the points are in linear relationship approximately. So we are going to use the least square fit to estimate a straight line $y = \beta_1 x + \beta_2$.

Before the least square fitting, there are some preprocessing work need to be done.

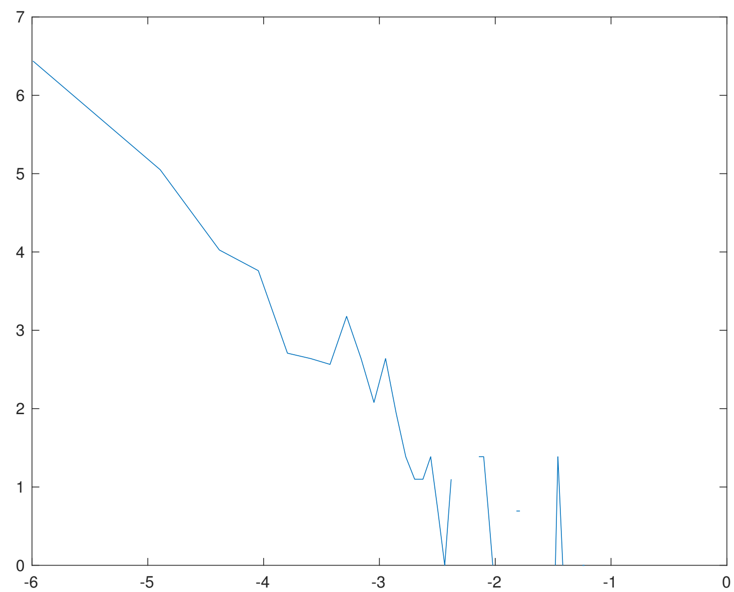


Figure 2.6: Line plot of double logarithm.

Take a look at the Fig 2.6, we can observe there are several disorders at the end of the plot. Go back to the distribution of laminar phases, in the last few bins, the number of laminar phases are unstable and small. In this specific case, from the 30th to the last bin, most bins have zero laminar phase in it and in other bins, a number of laminar phases are in the single unit. To have a deeper understanding of the distribution of laminar phase, we can take a look at the cumulative distribution function(cdf) of laminar phases.

To get the cdf, the laminar phases are ordered in ascending. In the case of threshold value 0.15, the total of laminar phases is 1043. So the probability of each laminar phases is 1/1043. A point (x,y) in cdf means the probability of a sample smaller than x is y. Like Fig 2.7 shows, the probability of a laminar phase smaller than

0.153 is 0.9789. On the contrary, there are $1043 \times (1 - 0.9789) \approx 22.0073$. In a general way, these insignificant samples can be removed.

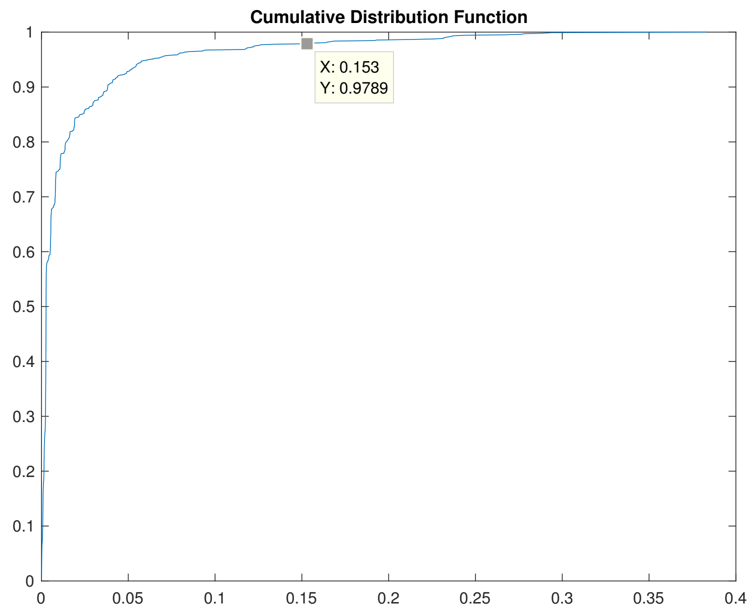


Figure 2.7: The empirical cumulative distribution function of laminar phases with threshold 0.15. The specified point (0.153, 0.9789), it means the probability of a laminar phase which is smaller than 0.153 is 0.9789

Besides, taking the fact that cannot take the logarithm of zero into account, we can find the first bin with zero laminar phases in it. In this case, the 20th bin is the first bin with zero laminar phase. So we only keep the first 19 bins for the double logarithm.

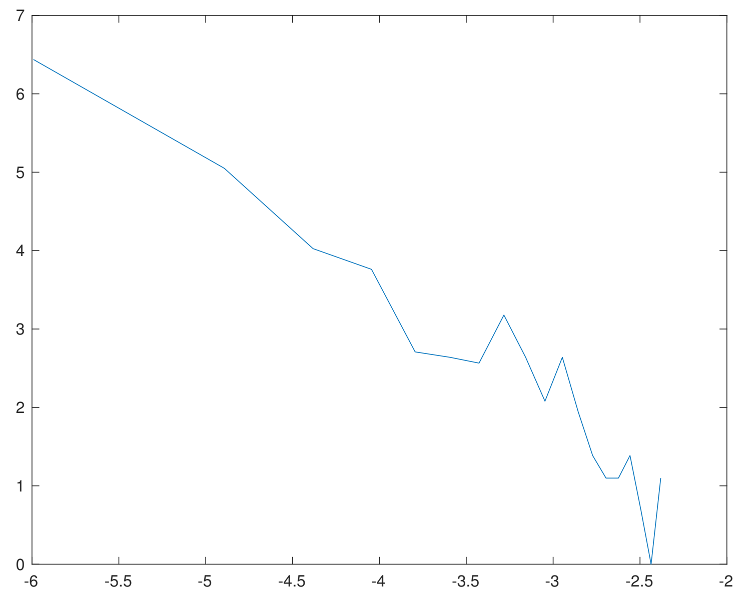


Figure 2.8: Line plot of double logarithm after preprocessing

From Fig 2.8, we can see that, after preprocessing, the line of the double logarithm is more likely to be a straight line. Then we are going to use the least square fitting method to estimate the ideal straight line.

In linear square fit method, with the best fit parameters, the squared difference between real value and modeled value is smallest. In current case, the $Ln(T)$ can be viewed as $(x_1, x_2, ..., x_{19})$ and the $Ln(P(T))$ is $(y_1, y_2, ..., y_{19})$. The hypothetical straight line is $y = \beta_1 x + \beta_2$. We would like to find the parameters which make the sum smallest:

$$S(\beta_1, \beta_2) = [y_1 - (\beta_1 x_1 + \beta_2)]^2 + [y_2 - (\beta_1 x_2 + \beta_2)]^2 + ... + [y_{19} - (\beta_1 x_{19} + \beta_2)]^2$$

To find the minimum of the sum, we can do the partial differential:

$$\frac{\partial S}{\partial \beta_1} = 0$$

$$\frac{\partial S}{\partial \beta_2} = 0$$

In this way, the parameters can be calculated. Finally the fitted line is $y = -1.6013x - 2.8978$, as displayed in Fig 2.9.

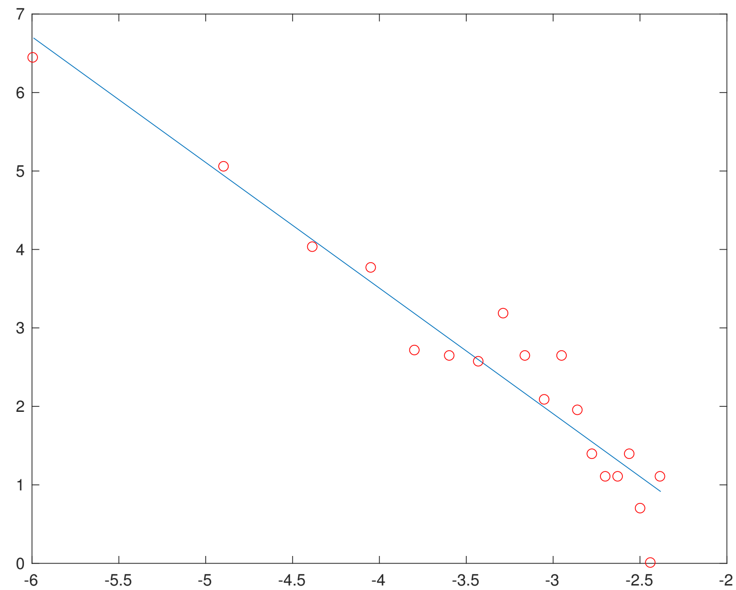


Figure 2.9: The straight line is: $y = -1.6013x - 2.8978$ with the $Ln(T)$ and $Ln(P(T)$ points.

It is worth reminding that this is only for the case with threshold value 0.15 and bin width 0.005. We need to verify the universality in other cases.

We can generate different distributions of laminar phases with different threshold values. Firstly we attempt to decrease the threshold value. If the threshold value is set to 0.12, there are 1552 laminar phases can be got from the log return time series. After the preprocessing work, there are 20 bins with 0.005 bin width remained for the next step. With the least square fit method, the parameters are: $\beta_1 = -1.7458, \beta_2 = -3.2946$. The straight line is displayed in Fig 2.9.

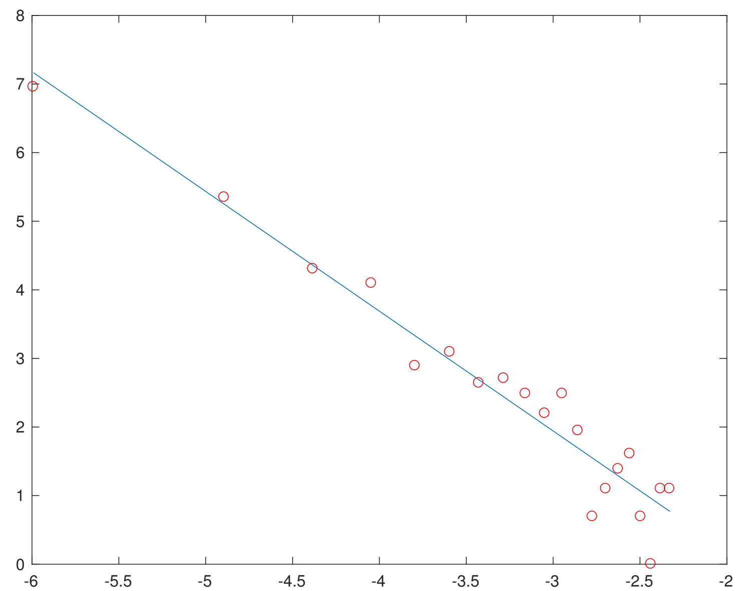


Figure 2.10: The fitting condition for case with threshold 0.12 and the bin width 0.005.

We continue to verify cases with a larger threshold value. In the case with threshold 0.2, there are 643 laminar phases generated in the log returns time series. To ensure the number of samples in double logarithm, the bin width is set to 0.01. After preprocessing work, there are 17 bins remained for the least square fit. Finally the estimated parameters are: $\beta_1 = -1.5142, \beta_2 = -2.0353$. The fitting condition is displayed in Fig 2.10.

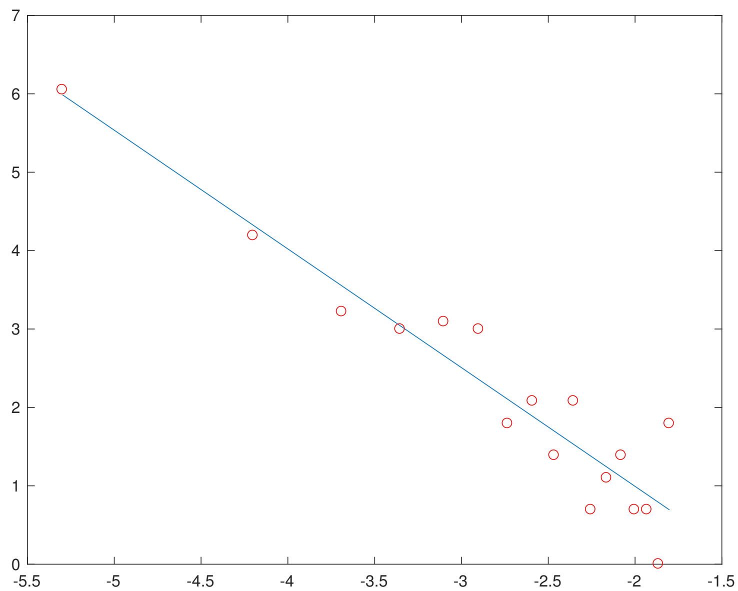


Figure 2.11: Fitting condition for the case with 0.20 threshold and 0.01 bin width

After these experiments on different cases, in conclusion, the estimated slopes β_1 are from -1.5 to -1.7 and the drifts are around -3. With the tunability of threshold values and bin width in histograms, it is possible to say that the slope can be closer to the theoretical value: -1.5. With this statement, we can say that the asymptotic -3/2 power law also can be detected in the intermittent behavior in Nord Pool log return.

Chapter 3

Conclusions

In this project, we analyze the log return time series of Nord Pool electricity prices and detect the existence and features of on-off intermittency in it.

Nord Pool electrical energy power price is distinguished with other kinds of financial data because of its high volatility. Intermittent behavior can be observed from its log return time series. Unlike other financial time series, the log return of Nord Pool prices is more stationary. So mean is subtracted from every sample as data preprocessing work.

To analyze the distribution of laminar phases, a threshold is needed to set at first. In the histogram of laminar phases, some bins are empty or only have several laminar phases in it. With the help of cdf, we know that removing samples of slight possibility may not have much effect on the following analysis. So we remove bins which contain single unit or zero laminar phases as preprocessing work in case of disorders.

To verify the theoretically asymptotic -1.5 power law of distribution of laminar phases, double logarithm is more convenient for estimating the slope and drift. With the least square fitting method, in different cases, the estimated slopes are from -1.7 to -1.5 and the drifts are around the range -2 to -3.

In conclusion, we can say that on-off intermittency exhibits in the Nord Pool prices log return time series. The distribution of laminar phases obeys the asymptotic -1.5 power law like the on-off intermittency in other signals.

Bibliography

[1] Bystrom, Hans NE. *Extreme value theory and extremely large electricity price changes.* International Review of Economics Finance 14.1 (2005): 41-55.

[2] Erzgraber, Hartmut, et al, *Time series analysis and long range correlations of Nordic spot electricity market data.*, Physica A: Statistical Mechanics and its Applications 387.26 (2008): 6567-6574.

[3] Grebogi, Celso, et al. *Final state sensitivity: an obstruction to predictability.* Physics Letters A 99.9 (1983): 415-418.

[4] Hammer, Philip W., et al. *Experimental observation of on-off intermittency.* Physical review letters 73.8 (1994): 1095.

[5] Heagy, J. F., N. Platt, and S. M. Hammel. *Characterization of on-off intermittency.*, Physical Review E 49.2 (1994): 1140.

[6] Kristiansen, Tarjei. *Forecasting Nord Pool day-ahead prices with an autoregressive model.* Energy Policy 49 (2012): 328-332.

[7] Mantegna, Rosario N., and H. Eugene Stanley. *Scaling behavior in the dynamics of an economic index.* Nature 376.6535 (1995): 46-49.

[8] Perell, Josep, et al. *Entropy of the Nordic electricity market: anomalous scaling, spikes, and mean-reversion.* Journal of Statistical Mechanics: Theory and Experiment 2006.11 (2006): P11011.

[9] Platt, N. S. E. A., E. A. Spiegel, and C. Tresser. *On-off intermittency: A mechanism for bursting.* Physical Review Letters 70.3 (1993): 279.

[10] Pomeau, Yves, and Paul Manneville. *Intermittent transition to turbulence in dissipative dynamical systems.* Communications in Mathematical Physics 74.2 (1980): 189-197.

[11] Tsay, Ruey S. *Analysis of financial time series.* Vol. 543. John Wiley Sons, 2005.

[12] Ufldl.stanford.edu. (2017). Data Preprocessing - Ufldl. [online] Available at:http://ufldl.stanford.edu/wiki/index.php/Data_Preprocessing#Per-example_mean_subtraction [Accessed 28 Aug. 2017].

[13] Weron, Aleksander, Szymon Mercik, and Rafal Weron. *Origins of the scaling behavior in the dynamics of financial data.* Physica A: Statistical Mechanics and its Applications 264.3 (1999): 562-569.

[14] Weron, Rafal. *Energy price risk management.* Physica A: Statistical Mechanics and its Applications 285.1 (2000): 127-134.

[15] Weron, Rafa, Michael Bierbrauer, and Stefan Trck. *Modeling electricity prices: jump diffusion and regime switching.* Physica A: Statistical Mechanics and its Applications 336.1 (2004): 39-48.

[16] Weron, Rafa, Ingve Simonsen, and Piotr Wilman. *Modeling highly volatile and seasonal markets: evidence from the Nord Pool electricity market.* The application of econophysics. Springer, Tokyo, 2004. 182-191.

[17] Weron, Rafa, and Adam Misiorek. *Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models.* International journal of forecasting 24.4 (2008): 744-763.

[18] Weron, Rafal. *Forecasting wholesale electricity prices: A review of time series models.* (2008).

[19] Weron, Rafa. *Market price of risk implied by Asian-style electricity options and futures.* Energy Economics 30.3 (2008): 1098-1115.