CONFIDENCE INTERVALS FOR CONDITIONAL EXPECTATIONS IN FINANCIAL TIME SERIES

GENC BERISHA

MTH6138

Dr. Wolfram Just

Queen Mary University of London

April 2020

Table of Content

List of Tables

List of Figures

CHAPTER ONE: INTRODUCTION

1.1 Background Theory

1.1.1 Nord Pool Power Market

Significant changes in the Nord pool power market are traced back to the 1990s, following new energy acts. Countries in the Nordic region pushed for established power trading and electricity production to market competition (Mundaca et al., 2013). As a result, all of the markets within the Nordic countries were liberalized, with Norway becoming the first country to liberalize its' market. Poor utilization of the resources which frequently led to overcapacity in the systems drove the reform for liberalization. Therefore, liberalization of the markets aimed to improve and facilitate the proper utilization of resources for production as well as the network transmission operation (Mundaca et al., 2013). Following the liberalization and integration of the markets, other regions with similar market objectives and policies are argued to have paid substantial attention to the Nordic region.

Over the years, the Nordic power market has been viewed to be functioning well, with proper liquidity in the primary products leading to general trust in terms of its transparency and efficiency (Zakeri and Syri, 2016). Several research institutions have also ranked the Nordic pool power market to be having the highest turnovers in exchange, particularly in Europe. Currently, the cross-border power market is the leading and most harmonized in the world (Hjalmarsson, 2000).

1.1.2 Time series

Time series refers to a sequential arrangement of data points that is the arrangement of data in agreement to their season of event (Hannan, 2009). In this context, time can be in units of hours, days, months, or even years, and is only a way one can relate the whole marvel to appropriate reference points.

1.1.1.1 Components for time series analysis

Time series components refer to the forces which influence the estimations of perception in a time series (Amini et al., 2016). Time arrangement segments allude to the powers which influence the estimations of perception in a period arrangement (Amini et al., 2016). Time series are of four sections; cyclic variations, random or irregular developments, seasonal variations, and trend. Seasonal and cyclic variations are the intermittent changes or transient vacillations (Dunsmuir and He, 2016).The trend shows the general inclination of the data, either increasing or decreasing pattern during an extensive stretch of time. A trend is a smooth, general, long haul, normal propensity (Rojas, 2018). It is not always a requirement that the increase or decrease follows the same direction throughout the given time frame.

There is another factor that causes the variation in the variable under examination. They are not normal variations but rather simply arbitrary or sporadic. These variances are unanticipated, uncontrollable, unpredictable, and erratic (Zumbach, 2015). They include floods, pandemics, wars, political instabilities, business sanctions, famines, and any other natural disasters.

1.1.1.2 Applications of time series analysis

Time series find applications in a wide range of disciplines. In making an investment decision, a time series tracks the development of the relevant data points, such as shares' and securities' price, over a predetermined time-frame with data points recorded at customary intervals (Amini et al., 2016). There is no base or upper limit in the amount of time that must be incorporated. This gives the researcher or analyst the freedom to collect enough data to give the required information (Diop and Kengne, 2017). Any variable that changes with time qualifies for time series analysis. In investment, for instance, time series is commonly used to track and predict the shares and securities prices over time. This tracking time varies from maybe an hour on a business day to the closing of the month over say five years (Kim and Jeong, 2015). This has proved to aid the investors in making informed decisions on when to buy or sell their shares, hence maximizing the profits.

Time series can be applied to evaluate how a given asset, economic, or security variable changes in a given period (Neuhierl and Weber, 2016). Similarly, it can also be used to investigate the variation of the chosen data points with other variables of interest over the given time duration (Zumbach, 2015). For instance, analyzing a daily time series of closing stock prices for a given stock requires one to obtain a chronological list of all the daily closing prices of the same stock for the past few years. This forms an annual, daily closing price time series for the stock. This time-series data can vary in several ways, either depending on seasons or economic variables (Rojas, 2018). In the case of seasonality variance, peaks and troughs may be observed at regular intervals each year. These peaks and troughs can be associated with particular seasons in the year, such as holiday, summer, or winter seasons. It is also possible to variation of the stock prices with natural variables such as pandemics or economic variables such as unemployment rate. Such a scenario exhibit dependency between stock price and the chosen economic or natural variables.

1.1.1.3 Time series forecasting

In time series forecasting, the future activity is predicted on the available information. This information is derived from historical data regarding the patterns associated with what is being predicted. In most cases, this relates to pattern analysis, repetitive change analysis, and seasonality issues. Just like all other forecasting methods, time series forecasting does not guarantee success but only higher chances.

Time series is also instrumental in load forecasting in electrical power systems (Louie, 2017). Just like in the stock price prediction, electrical power companies collect data about the power demands, equipment repairs, and the availability of raw materials (such as water, wind or coal) used in power generation for a given period. They use this data to predict power demands in the future as well as scheduled maintenance for equipment. This helps in planning, sustainability, and profit maximization (Kim and Jeong, 2015).

In this project, confidence intervals for conditional expectations in financial time series are investigated. The data is the Nord Pool spot market from the Nordic electricity market.

1.2 Objectives

The primary objective of this project is to analyze whether there is a visible impact of the daily variations in power demand on the returns of the spot market price by computing the relevant confidence intervals.

1.3 Significance of the Study

The project investigates whether the hourly variations of electricity price affects Nord pool power returns. The findings of this project are not only useful to electrical power companies but also other firms whose product or service cost varies with time. In predicting the effects of hourly cost variations, the findings of this project will be an important tool that can be used by Nord power in evaluating whether it will accrue more benefits when the price volatility is minimized.

CHAPTER TWO: METHODOLOGY

This chapter describes the approach used to achieve the project objectives. The software and tools used are described, with an attempt to justify their choice over other available options. The description of the data used is also given, the analysis approach and the formulas for deriving the required variables.

2.1 Software and Tools Used

In this project, Python programming language is used to plot and analyses the time series of the given data and their confidence intervals. The version of Python used is Python 3.7, running on the Jupyter Notebook environment. The choice of Python was informed by its acceptance as one of the main tools used in statistics and data analysis (Kumar, 2019). This is made possible by the many open source packages available for such applications. Among these packages are Pandas, Numpy, Statsmodels, Matplotlib, and Seaborn. Jupyter Notebook environment is chosen because it allows interactive coding, whereby a code segment can be executed and the results visualized or modified if need be before proceeding to the next section (Perkel, 2018). This is particularly useful in statistics and data analysis, where exploratory data analysis and data munging is required before the actual analysis start. In such situations, results for both before and after data, munging need to be shown, together with the source code generating such results, and in a procedural way for comparison (Yakimchik, 2019). Jupyter Notebook environment also allows people to publish and share their code in various formats such as PDF, HTML, and Python scripts. Such flexibility will allow even people who might not be having the environment set up to run such code follow the flow, with both source code and the results (Perkel, 2018).

2.2 Meet the Data

The data provided is the Nord Pool Power Market data as recorded from 01/01/99 to 26/01/07. The figures are recorded on an hourly basis between 12 am to 11 pm, specifically on electricity prices during that particular hour. From this data, the returns on electricity sale are calculated from the formula:

$r(t) = ln(\frac{s(t)}{s(t-1)})$ ..... *(3.1)*, where:

r (t) are the returns as a function of time,

s (t) is the current hour price, and

s (t-1) is the previous hour price.

Since the data provided is vast, containing more than 70,000 rows, Microsoft Excel is used to compute the returns r (t) from the formula given in equation (3.1) before the analysis with Python commences.

2.3 Data Analysis

The data analysis part was started by visualizing the data and to find if there was any missing value. Abelairas and Astorkiza (2020) highlight the importance of exploratory data analysis. According to these researchers, exploratory data analysis is a critical component in any data analysis task since besides giving an overall view of data; it also helps in identifying unusual cases and extreme values, as well as the obvious errors in the data. Obvious errors may include missing values or incorrect data types (Abelairas-Etxebarria and Astorkiza, 2020).

Only one value in the Returns column was found to be missing. This was expected since looking at the formula in equation (3.1), the returns are a function of the electricity price for both current and previous. When the recording starts, the information on the previous hour's price is not available. As a result, the first row will have no valid figure for the returns. Hence it is left null. And indeed, looking at the excel data, the first record has no value on the returns. Since one record is negligible compared to over 70,000 other records, the row with the missing value can be dropped without appreciably distorting the data (Albers and Gower, 2010). So this row is dropped.

Next, the Date column in the data is analyzed. The purpose of focusing on this column was to determine if indeed the data was recorded in every hour for the period specified, as well as observe the general trend in the prices recorded. If this is found to be true, then applying a time series analysis for data observations taken at consistent intervals in time will be appropriate (Perron and Zorita, 2017). To accomplish this, we first confirm that the Date column is cast to Python DateTime data type since this is a requirement (Campos-Rozo and Domínguez, 2016). We group the data by month and plot the records count every month. This data is then decomposed into time series components using Python Statsmodels package to give the variations, trend, and seasonality (Perron and Zorita, 2017). These trends reveal that the data was indeed recorded for all 24 hours a day throughout the period stipulated. This is ascertained by the uniform variations in the seasonality, repeating each year, and constant trends in all years, except the leap years 2,000 and 2,004, which shows more records than other years, due to the extra one day in a leap year.

There are two broad modes of time series analysis: time-domain and frequency-domain analysis (Perron and Zorita, 2017). Time-series analysis methods have their origin in the classical theory of correlation and principally deal with covariance and cross-covariance functions. The resulting models are of auto-regressive, moving-average type, mostly for single series (Fokianos and Promponas, 2011). Such models can be described as sophisticated variants of the linear regression model. Frequency-domain methods, on the other hand, entail spectral analysis. They extend Fourier analysis, which approximates analytic functions by taking a weighted sum of sine and cosine functions of harmonically increasing frequencies (Fokianos and Promponas, 2011). Since data analyzed in this project is relatively simple and contains an only one-time variable, time-domain methods are used. Satisfied with the findings, time series plots are made for both the hourly price and the returns. The plots are made first using matplotlib package, then with seaborn package with the confidence intervals indicated.

CHAPTER THREE: RESULTS

This chapter presents the results obtained by following the methodology procedure, as outlined in Chapter Two. It consists of mostly plot figures and Python code snippets outlining how the plots came about. The code snippets are not chronological in a way that they can be consolidated together to give the desired plots.

3.1 Reading the Excel Data file and Visualizing the Data

```
# Read Excel Data file
nord_pool = pd.read_excel("Data.xlsx")
# Cast the 'Date' column to datetime format
nord_pool['Date'] = pd.to_datetime(nord_pool['Date'], infer_datetime_format=True)
# Describe data
nord_pool.describe()
```

Table 1: Data Description

	Price	**Returns**
count	70752.000000	70751.000000
mean	27.475522	0.000007
std	14.713680	0.059092
min	2.330000	-1.263635
25%	16.750000	-0.016477
50%	25.880000	-0.002933
75%	32.790000	0.012434
max	238.010000	1.954708

The code for removing the row with some missing value is omitted but will be found in the complete script.

```
import datetime
# Convert 'Date' to datetime
```

```
fixed_dates_df = nord_pool_date.copy()

fixed_dates_df["Date"] = fixed_dates_df["Date"].apply(pd.to_datetime)

fixed_dates_df = fixed_dates_df.set_index(fixed_dates_df["Date"])


# Group by month

grouped = fixed_dates_df.resample("M").count()

data_df = pd.DataFrame({"count": grouped.values.flatten()}, index=grouped.index)

data_df.head(10)
```

Table 2: First 10 rows of the Date grouped by Month

Date	Count
1999-01-31	743
1999-02-28	672
1999-03-31	744
1999-04-30	720
1999-05-31	744
1999-06-30	720
1999-07-31	744
1999-08-31	744
1999-09-30	720
1999-10-31	744

```
# Plot the records count by month

from pandas.plotting import register_matplotlib_converters

register_matplotlib_converters()

plt.style.use("ggplot")

rc = data_df.plot(color="#a50656")

rc.set(xlabel="Time (In Years span)", ylabel = "Record Count")

fig = rc.get_figure()
```

```
fig.savefig('Record_Count.png')
```

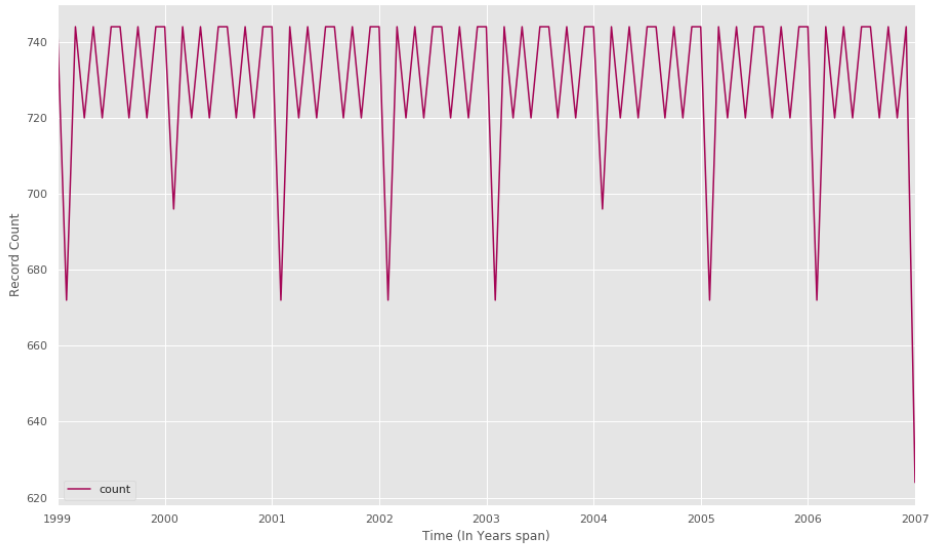

Figure 1: Monthly Record Count Plot

```
# Time Series Components of the Monthly Record Count

from statsmodels.tsa.seasonal import seasonal_decompose

result = seasonal_decompose(data_df)

ts = result.plot()

ts.savefig('TS_Components.png')
```

Figure 2: Time Series Components of the Monthly Record Count

3.2 Time Series Plot

```
# Plot Price Time Series

plt.xlabel("Time (In Years span)")

plt.ylabel("Electricity Price (EUR/MWh)")

plt.plot(nord_pool['Date'], nord_pool['Price'], color="#ca08bb")

plt.savefig('Price.png')
```

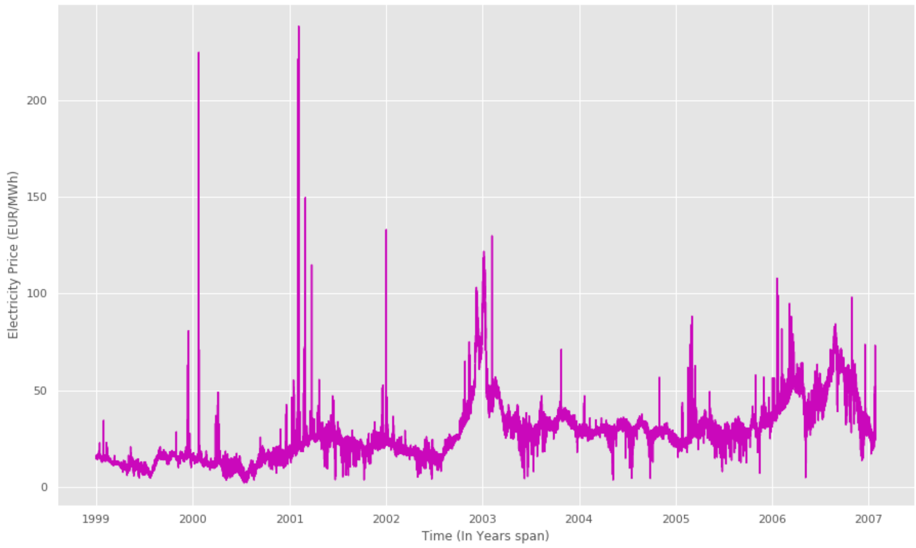

Figure 3: Hourly Price Time Series Plot

```
# Plot Returns Time Series
plt.xlabel("Time (In Years span)")
plt.ylabel("Returns")
plt.plot(nord_pool['Date'], nord_pool['Returns'], color="#ca08bb")
plt.savefig('Returns.png')
```

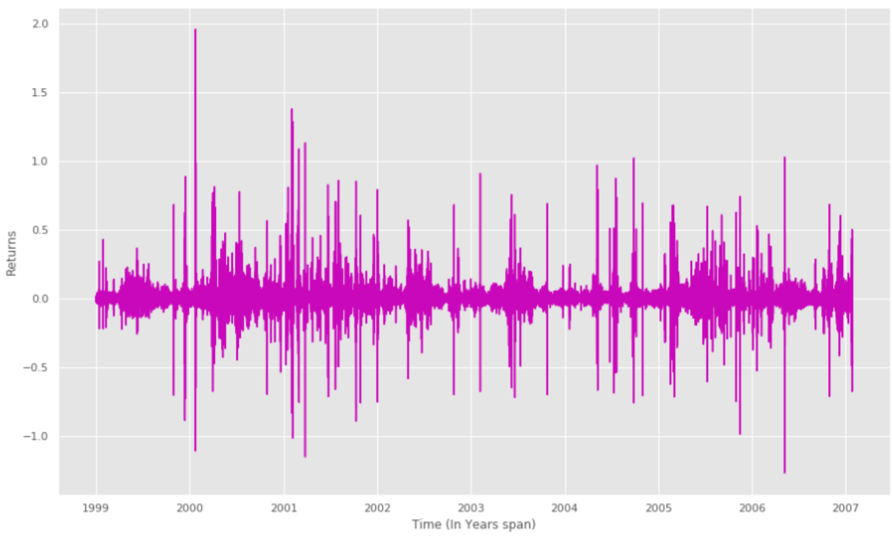

Figure 4: Hourly Logarithmic Returns Time Series Plot

Price Time Series with Confidence Interval

price = sns.lineplot(x="Date", y="Price", color='#ca08bb', data=nord_pool)

price.set(xlabel="Time (In Years span)", ylabel = "Electricity Price (EUR/MWh)")

fig = price.get_figure()

fig.savefig('Price_CI.png')

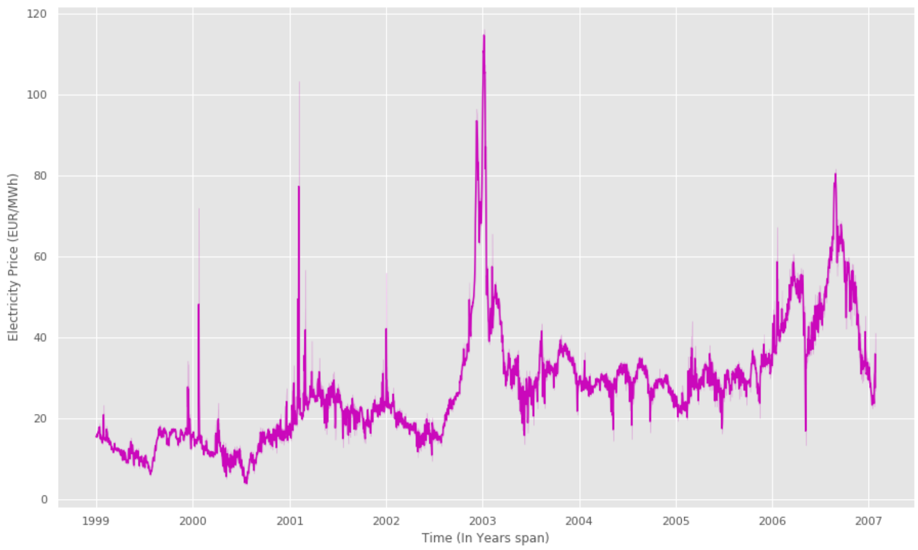

Figure 5: Hourly Price Time Series Plot with Confidence Intervals

Returns Time Series with Confidence Interval

returns = sns.lineplot(x="Date", y="Returns", color='#ca08bb', data=nord_pool)

returns.set(xlabel="Time (In Years span)", ylabel = "Returns")

fig = returns.get_figure()

fig.savefig('Returns_CI.png')

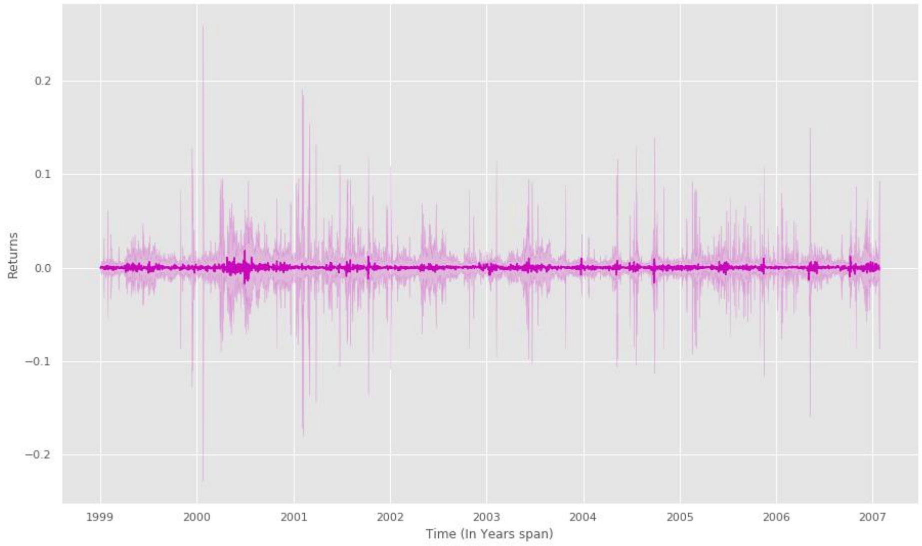

Figure 6: Hourly Logarithmic Returns Time Series Plot with Confidence Intervals

CHAPTER FOUR: ANALYSIS AND DISCUSSION

This chapter analyses the results obtained in chapter three based on their relevance in comparison to the findings by other studies.

4.1 Analysis of the Findings

From Figure 3 and Figure 4, there seems to be a decline in electricity prices in mid-1999, which results in a slight increase in the returns. The price surges again but for a short duration in late 1999 and early 2000, which results in a corresponding spike-like surge in returns in the same period. There is again a drop in the electricity price in mid-2000, which results in a considerable increase in returns. The price increases again in late2000 through to early 2001, but this time the higher prices seem to persist a bit longer. The result is seen as a consistently higher return in this period. Such a phenomenon occurs again in late 2002 through to early 2003, but this time they seem to fluctuate throughout this period, with low returns taking dominance. The same trend in both electricity prices and corresponding returns is observed in early and late 2006.

More to the above specific observations, there seems to be a general decline trend in electricity prices every mid-year with the exception of 2004 and 2006. Such a decline in most cases results in a considerable increase in the returns from the electricity sale. There also seems to be a general growth in the electricity price near the end of the year, which runs through the first quarter of the following year. This increase in most cases does not result in a corresponding increase in the returns, though. These observations are made from the general time series plot of the respective parameters (price and returns) without bothering about the validity of the results realized; that is, the confidence interval the assertion of the results.

Looking at the plots in Figure 5 which takes into account the confidence interval for the electricity price, it is observed that the moving average closely follows the plot in Figure 3, except in cases where there exist very price increase lasting for a concise duration, such as in early 2000 and 2001. This was expected since the price variations are recorded as used as linear values. This plot is not of much interest in the remaining part of this project. It just verifies to us that the electricity prices indeed vary over time.

The returns time series plot with confidence intervals in Figure 6 is what is of primary importance to us since it has the answers to our main objective. This plot exhibits a wide disparity in the hourly logarithmic returns with the average mean, both in values and in trend. For instance, even though the same plot without the confidence interval shows logarithmic returns variations in excess of +/-0.2, the average values rarely surpass +/-0.01. In terms of trend, the average changes in returns do to necessarily follow the general change. Such instances are observed in mid-1999, where despite a constant boom depicted by the general plot, the average plot shows a recession just in the middle of this period. The same case is observed in early 2001. There are also instances where the average logarithmic returns show an increasing trend while the general trend is declining, such as in early 2007. This suggests that other than the electricity prices, there are other factors that affect the returns. Since the hourly returns are on a logarithmic scale, even a slight change is significant and should not be assumed.

4.2 Discussion

The preceding analysis reveals that the fluctuations in electricity price in the Nordic spot electricity market has no direct effects on the returns the company gets. These findings are in agreement with the findings by Voronin, Partanen, and Kauranne (2013), who studied the same market and Grimm and Zoettl (2013), who studied the general electricity market. Even though there exists a notable variance in both the electricity prices and the hourly logarithmic returns, the two might be influenced by various factors such as the entrance of the new customers in the market, availability of the raw materials required to produce electricity and the tendency of customers resolving to alternative sources of energy (Mauritzen, 2013).

The logarithmic model used in computing the returns could also be a source of error. The justification of its adoption and the accuracy to which it estimates the returns is not justified in any way (Grimm and Zoettl, 2013). Additionally, since it is based on a logarithmic scale, it might diminish even significant variations and make them look negligible at a glance.

CHAPTER FIVE: CONCLUSION

The primary objective of this project was to find out whether the hourly variation of electricity for the Nordic electricity spot market affects the revenue returns from the electricity sale. To achieve this, a time series analysis was done, and the confidence interval of the hourly logarithmic returns was used to investigate the effect of price variations on the returns. The findings reveal that even though there are significant variations in both the electricity price and the returns, the trend in return variations cannot be absolutely attributed to the price trend. This is so because the two curves do not seem to follow each other in either their periodic or irregular trends.

To start with, the electricity prices seem to be highest towards the start and end of every year and drop to the lowest values around the mid of the year. This can be attributed to seasonal variations in the price of generating electricity since this is what determines, to a great extent, the power unit price. During the start and the end of the year seasons, there is a likelihood that the cheaper means of generating electricity are not available. Such could be caused by insufficient water in rivers and reservoirs for a hydroelectric generation or slow wind speed for wind power generation. Such occurrences might force the power generating companies to resolve to other means such as the use of coal and diesel engines, which are relatively expensive. The extra cost will be passed to customers in the form of higher electricity prices to sustain the company. When such happens, some customers are likely to resolve to alternative means of energy such as solar panels, reducing the Nordic customer base. This will eventually result in low returns.

Around mid-year, there could be two possibilities. One of the possibilities is that the cheaper means of generating electricity is available, such as enough water in rivers and reservoirs for hydroelectric power generation. This translates to low electricity prices, and most people will now switch to the main grid as their primary source of power. This increases the customer base, which translates to more units of power being sold and hence higher returns. Another possibility could be that due to winters, people will need electrical power to keep their houses warm. The solar energy may not be available at such a season, and even if it was available, it is not reliable for warming houses. The government could, therefore, provide a mandatory subsidized in electricity price since it is a necessity. This will result in low prices, as observed in the price curve. With almost everybody using the Nordic electric power, the customer base is large, translating to more returns.

It is also observed that electricity price variations do not affect the returns in a predictable pattern. This suggests that there are other factors, apart from the price, which is not accounted for (at least in this) project. Such factors might include the entry of new competitors or customers in the market. The Nordic company should consider conducting market research to determine these other possible factors. After they are identified, a similar analysis should be carried on them to determine the extent to which they affect the returns. Doing so will aid in devising means to maximize the returns since the company will be in a position to make informed decisions and mitigate appreciable factors while ignoring the minor ones.

Low returns in any company are undesirable. The hike in electricity prices seems to result in a drop in returns for the Nordic electricity market spot. The company should, therefore, consider devising a mechanism to mitigate this trend in the affected period. This applies to other companies as well. The cost of producing any commodity, as well as the demand for that particular, varies with seasons (Grimm and Zoettl, 2013). Identifying these seasons and the underlying factors behind these variations are key to business success as it will help in planning and providing possible losses which can be associated with such variations.

References

Albers, C.J., and Gower, J.C., 2010. A general approach to handling missing values in Procrustes analysis. *Advances in Data Analysis and Classification*, **4**(4), pp.223–237.

Amini, M.H., Kargarian, A. and Karabasoglu, O., 2016. ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. *Electric Power Systems Research*, **140**, pp.378–390.

Campos-Rozo, J.I. and Domínguez, S.V., 2016. A Python-based interface to examine motions in time series of solar images. *Proceedings of the International Astronomical Union*, **12**(S327), pp.25–27.

Diop, M.L. and Kengne, W., 2017. Testing parameter change in general integer-valued time series. *Journal of Time Series Analysis*, **38**(6), pp.880–894.

Dunsmuir, W. and He, J., 2016. Marginal estimation of parameter driven binomial time series models. *Journal of Time Series Analysis*, **38**(1), pp.120–144.

Fokianos, K. and Promponas, V.J., 2011. Biological applications of time series frequency domain clustering. *Journal of Time Series Analysis*, **33**(5), pp.744–756.

Grimm, V. and Zoettl, G., 2013. Investment incentives and electricity spot market competition. *Journal of Economics & Management Strategy*, **22**(4), pp.832–851.

Kim, Y.-J. and Jeong, M.-A., 2015. Efficient anomaly detection through confidence interval estimation based on time series analysis. *International Journal of Advanced Smart Convergence*, **4**(2), pp.46–53.

Kumar, N. and Gaurav, 2019. Data analysis and data visualization using Python. *International Journal of Computer Sciences and Engineering*, **7**(5), pp.1287–1291.

Louie, H.M., 2017. Time-series modeling of aggregated electric vehicle charging station load. *Electric Power Components and Systems*, **45**(14), pp.1498–1511.

.

Mauritzen, J., 2013. Dead battery? Wind power, the spot market, and hydropower interaction in the Nordic electricity market. *The Energy Journal*, **34**(1).

Neuhierl, A. and Weber, M., 2016. Monetary policy and the stock market: time-series evidence. *SSRN Electronic Journal*.

Perkel, J.M., 2018. Why Jupyter is data scientists' computational notebook of choice. *Nature*, **563**(7729), pp.145–146.

Perron, P. and Zorita, E., 2017. Time series methods applied to climate change. *Journal of Time Series Analysis*, **38**(5), pp.639–639.

Rojas, I., 2018. Forecasting: a new open access journal dealing with time series analysis and forecasting. *Forecasting*, **1**(1), p.1.

Voronin, S., Partanen, J. and Kauranne, T., 2013. A hybrid electricity price forecasting model for the Nordic electricity spot market. *International Transactions on Electrical Energy Systems*, **24**(5), pp.736–760.

Yakimchik, A.I., 2019. Jupyter Notebook: a system for interactive scientific computing. *Geofizicheskiy Zhurnal*, **41**(2), p.121.

Zumbach, G., 2015. Cross-sectional universalities in financial time series. *Quantitative Finance*, **15**(12), pp.1901–1912.