
Testing the Assumptions of Normality and Independence of
Returns of the Black-Scholes Option Pricing Formula

Research Question:

How well do the Oslo, 2018, power prices satisfy the assumptions of normality and independence of the Black-Scholes option pricing formula?

Subject: Mathematics

Name: Simran Sahajpal

TABLE OF CONTENTS

1	Introduction.....	2
1.1	The Black-Scholes Formula and Risk-Neutral Pricing.....	2
1.1.1	Assumptions of the Black-Scholes Formula.....	3
1.2	The Normal Distribution.....	4
1.3	Power Market Data Sample.....	5
1.3.1	Time-series of Power Prices.....	5
2	Returns.....	6
2.1	Arithmetic Returns (AR).....	7
2.2	Geometric Returns (GR).....	7
2.3	Comparing Arithmetic Returns with Geometric Returns.....	8
3	Normality Testing of Arithmetic Returns (AR).....	13
3.1	Qualitative Testing.....	13
3.1.1	Visual Evaluation of Density Histogram.....	13
3.1.2	The Normal Q-Q Plot.....	16
3.2	Hypothesis Testing.....	19
3.2.1	Declaration of Hypotheses.....	19
3.2.2	Choice of Normality Tests.....	19
3.2.3	Kolmogorov-Smirnov Test (K-S).....	20
3.2.4	Anderson Darling Test (A-D).....	21
3.2.5	Shapiro-Wilk Test (S-W).....	21
4	Normality Testing of Geometric returns (GR).....	22
4.1	Qualitative Testing of Data.....	22
4.1.1	Visual Evaluation of Histogram.....	22
4.1.2	Normal Q-Q Plot.....	23
4.2	Hypothesis Testing.....	23
5	Independence of Returns.....	24
5.1	Covariance.....	25
5.2	Pearson's Correlation Coefficient.....	25
5.3	Autocorrelation of Returns.....	26
5.3.1	Correlation of Consecutive Returns.....	26
5.3.2	The Sample Autocorrelation Function.....	27
6	Conclusion.....	30
7	References.....	31
8	Appendix.....	33

1 INTRODUCTION

The statistical analysis and modelling of the patterns, complexity and stochasticity of power prices is essential to understand and forecast their behaviour. Many financial models and formulae are derived based on the behaviour of prices for uses such as option-pricing. Using an accurate and suitable model is critical to avoid unacceptable financial risks. This essay focuses on the core assumptions of the widely-used Black-Scholes mathematical formula for European call-option pricing to investigate the research question: *How well do the Oslo, 2018, power prices satisfy the assumptions of normality and independence in the Black-Scholes option-pricing formula?*

This report is divided into three parts: understanding the relevant assumptions of the Black-Scholes formula, evaluating the usage of the forms of returns and testing the assumptions against spot-price data from the Norwegian power market.

1.1 THE BLACK-SCHOLES FORMULA AND RISK-NEUTRAL PRICING

A European call option gives the right, but not the obligation to purchase an asset at a specified exercise price and time. The Black-Scholes formula returns a single price for the option that eliminates any opportunity for arbitrage, meaning that the price allows no risk-free profit for a trader and is therefore often called the risk-neutral valuation of options (Ross, 2006).

The price of a European call option C_0 is given by:

$$C_0 = S_0 N(d_1) - X e^{-rT} N(d_2)$$

Where S_0 is the current price, X and T are the exercise price and time respectively, r is the risk-free interest rate, and $N(x)$ is the standard normal cumulative distribution function (CDF).

$N(x)$ gives the probability for a random observation from the sample to be less than or equal to a certain value, x .

The inputs of $N(x)$ are given by,

$$d_1 = \frac{\ln\left(\frac{S_0}{X}\right) + \left(r + \frac{\sigma_r^2}{2}\right)T}{\sigma_r\sqrt{T}}$$

$$d_2 = \frac{\ln\left(\frac{S_0}{X}\right) + \left(r - \frac{\sigma_r^2}{2}\right)T}{\sigma_r\sqrt{T}}$$

Where, σ_r is the standard deviation of returns (Khan, 2013). Standard deviation, an indicator of market volatility, is given by:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2}$$

Where n is the sample size, x_i is a data point for $i = 1, 2, \dots, n$, and μ is the mean.

The probability, $N(d_1)$, acts as a weight to S_0 , while Xe^{-rT} gives the discounted exercise price. When σ_r is increasing, d_1 increases due to the $+\sigma_r^2$ in the numerator, while d_2 is decreasing due to the $-\sigma_r^2$ in then numerator. Hence, $N(d_1)$ increases while $N(d_2)$ decreases. Consequently, when σ_r is high, the value of the option is high. Therefore, increased volatility implies a higher option price. This makes sense as if the market is volatile, there would be a greater chance for the future price to be much higher than the strike price, as well as being much lower. A higher price would allow no risk-free profit.

1.1.1 Assumptions of the Black-Scholes Formula

The Black-Scholes formula assumes that the underlying prices of an asset follow geometric Brownian motion. This means that *returns follow a normal distribution* and that future price changes are independent of past price movements (Ross, 2006). The latter premise implies that *future returns are uncorrelated to past returns*, however, the validity

of it is often disputed. Investors in agreement argue that it is a consequence of the Efficient-Market Hypothesis (EMH) which states that prices reflect all available information and “no amount of analysis can give an investor an edge over other investors” (Thune, 2019). Those in disagreement argue that information is absorbed by investors at different rates, and thus, future price changes will tend to follow past price movements (Ross, 2006).

As seen, σ_T is a crucial parameter of the Black-Scholes formula in which the assumptions explained are contained. Incoherence with the assumptions may therefore, return an option price allowing arbitrage, hence dissatisfying its initial requirements. The remaining report is divided into exploring the two assumptions concerning returns:

1) Normality

2) Independence

1.2 THE NORMAL DISTRIBUTION

The normal distribution is recognised for its presence in many natural phenomena. The perfect symmetry around the mean characterises the distribution. Its probability density function (PDF) has two parameters: mean (μ) and standard deviation (σ) and is given by:

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

In a standard normal distribution (s.n.d), μ is 0 and σ is 1.

One property of the normal distribution that is used onwards is the 68-95-99.7 rule which gives the approximate probability of occurrence for three different intervals (Table 1).

Data intervals	Probability of Occurrence (%)
$\pm 1 \sigma$ of μ	68
$\pm 1.96 \sigma$ of μ	95
$\pm 3 \sigma$ of μ	99.7

Table 1: 68-95-99.7 rule

1.3 POWER MARKET DATA SAMPLE

To test if the power market meets the assumptions of the Black-Scholes formula, a data sample of 365 consecutive daily spot-prices/MWh of Oslo, Norway, 2018, was taken from Nord Pool Historical Market Data (Appendix A). Some considerations of the data include that daily-prices should be chosen over hourly-prices to avoid cyclic trends that arise due to the influence of cyclical demand patterns (Simonsen, 2004). The choice of year, 2018, is assumed to represent the population of power prices in Norway. The data size of 365 is also a sufficiently large sample size which is beneficial when conducting the normality tests ahead.

1.3.1 Time-series of Power Prices

Financial prices can be modelled by a discrete time-series, a set of observations measured at equidistant time (t) intervals. This means that for any day, n , $\Delta t = t_n - t_{n-1}$, is constant.

Figure 1 shows the data sample presented as a time-series, where no clear trend is seen.

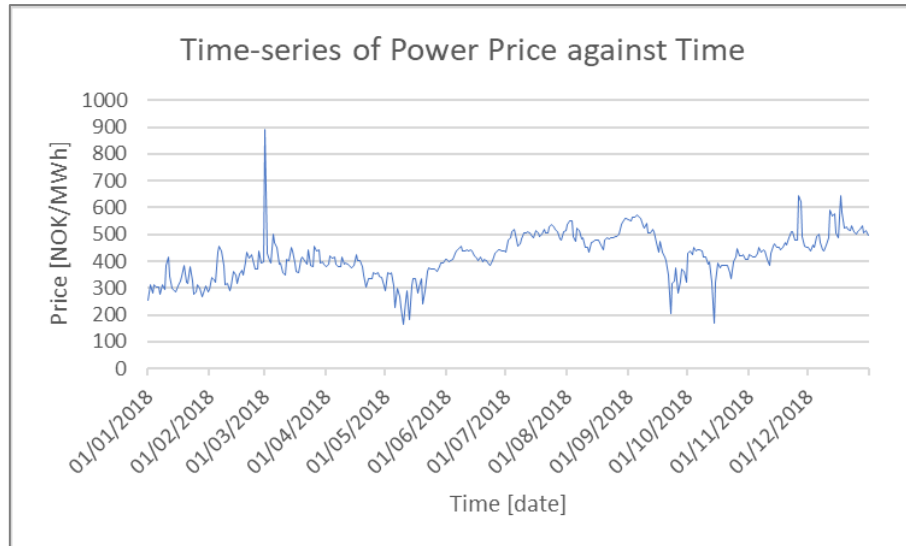


Figure 1: Time-series of Oslo, 2018 spot-prices

Before testing the assumptions, the forms of returns are explored.

2 RETURNS

Power prices are affected by many underlying factors, for instance, weather. The correlation between consecutive prices from the presence of trend and seasonality is often removed when working further with a time-series to transform it into stationary, leaving only the *short-run fluctuations*. In a stationary time-series, properties such as mean and variance (σ^2) are constant. Such statistical estimators of future behaviour are *only* useful if the series is stationary (Nau, n.d).

Differencing, $d(n)$, a transformation to make the price series stationary, yields the change in consecutive prices and is given by:

$$d(n) = p(n) - p(n - 1),$$

where n is a day, $p(n)$ is the corresponding end-of-day price, and the number of differenced values is $n-1$.

Concerning the Black-Scholes formula, **returns** are more effective measures of change as they comprise of a *relative change* in the value of an asset and the standard deviation of returns is an important parameter of the formula. Differencing does not reflect how significant the change is relative to the original price. Section 2.1 and 2.2 discusses the two forms of returns.

2.1 ARITHMETIC RETURNS (AR)

AR (r_t) is the ratio of the difference, $d(n)$, to the reference value, $p(n - 1)$.

$$r_t = \frac{d(n)}{p(n - 1)} = \frac{p(n) - p(n - 1)}{p(n - 1)} = \frac{\Delta p(n)}{p(n - 1)}$$

Therefore, a change in price $\Delta p(n)$, when $p(n - 1)$ is small, would result in a greater percentage return in comparison to if $p(n - 1)$ is large.

2.2 GEOMETRIC RETURNS (GR)

GR (R_t) is the difference in the logarithms of consecutive prices. Let,

$$L(n) = \ln(p(n)),$$

and,

$$R_t = L(n) - L(n - 1)$$

$$R_t = \ln(p(n)) - \ln(p(n - 1))$$

$$\therefore R_t = \ln\left(\frac{p(n)}{p(n-1)}\right)$$

Hence, a geometric return is the natural logarithm of the ratio of the price at day n to the price at day $n-1$.

2.3 COMPARING ARITHMETIC RETURNS WITH GEOMETRIC RETURNS

Both return forms are often used in financial analysis, and it was observed throughout several data values, that AR were approximately equal to GR. For instance, the arithmetic return of January 4th to January 5th is

$$\begin{aligned} &= \frac{304.33 - 311.56}{311.56} \\ &= \frac{-7.23}{311.56} \\ &= -0.0232 \text{ (4 d.p.)} \end{aligned}$$

And the geometric return is

$$\begin{aligned} &= \ln\left(\frac{304.33}{311.56}\right) \\ &= -0.0234 \text{ (4 d.p.)} \end{aligned}$$

Here, the difference between GR and AR was only $-0.0234 - (-0.0232) = 0.0002$, which seemed negligible.

It was, therefore, difficult to understand which return to use over the other. To further evaluate the difference, a Maclaurin series expansion was done on the GR, which is the Taylor series expansion centred about 0. This expands the function into an infinite sum of certain polynomials given by its higher derivatives that would approximate the GR in the proximity of 0. The first step was to rewrite GR as a function of AR.

$$\begin{aligned} \text{since, } R_t &= \ln\left(\frac{p(n)}{p(n-1)}\right) \\ R_t &= \ln\left(\frac{p(n) - p(n-1)}{p(n-1)} + \frac{p(n-1)}{p(n-1)}\right) \\ R_t &= \ln\left(\frac{p(n) - p(n-1)}{p(n-1)} + 1\right) \\ R_t &= \ln(r_t + 1) \end{aligned}$$

The function can be expressed using the Maclaurin power series expansion theorem:

$$\sum_{n=0}^{\infty} c_n x^n, \text{ where } c_n = \frac{f^{(n)}(0)}{n!}$$

Thus, the term c_n is given by the n th derivatives of a function $f(x)$ at $x=0$, divided by the n th factorial.

In this case, x is AR and $f(x)$ is GR,

$$\text{Hence, } f(x) = \ln(x + 1)$$

When $n=0$:

$$f(0) = \ln(1 + 0) = 0$$

$$c_0 = 0$$

When $n=1$:

$$f'(x) = \frac{1}{1+x}$$

$$f'(0) = 1$$

$$c_1 = \frac{1}{1!} = 1$$

When $n = 2$:

$$f''(x) = \frac{0(1+x) - 1(2)}{(1+x)^2}$$

$$f''(x) = \frac{-1}{(1+x)^2}$$

$$f''(0) = -\frac{1}{1} = -1$$

$$C_2 = -\frac{1}{2!}$$

$$C_2 = -\frac{1}{2}$$

When $n = 3$:

$$f'''(x) = \frac{0 + 1(2)(1+x)}{(1+x)^4}$$

$$f'''(x) = \frac{2(1+x)}{(1+x)^4} = \frac{2}{(1+x)^3}$$

$$f'''(0) = 2$$

$$C_3 = \frac{2}{3!} = \frac{2}{6} = \frac{1}{3}$$

When $n = 4$:

$$f^4(x) = \frac{0 - (2)(3)(1+x)^2}{(1+x)^6}$$

$$f^4(x) = \frac{-6(1+x)^2}{(1+x)^6}$$

$$f^4(x) = \frac{-6}{(1+x)^4}$$

$$f^4(0) = -6$$

$$C_4 = \frac{-6}{4!}$$

$$C_4 = \frac{-6}{24}$$

$$C_4 = -\frac{1}{4}$$

Thus,

$$f(x) = \ln(1+x) = C_0 + C_1x + C_2x^2 + C_3x^3 + C_4x^4 + \dots + C_nx^n$$

$$f(x) = \ln(1+x) = 0 + x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots$$

When $|x| \ll 1$, the first-order term of $f(x)$, x , dominates, as $|x^k| \rightarrow 0$, for $k = 2, 3, 4 \dots$

Thus, when $r_t \ll 1$,

$$\ln\left(\frac{p(n)}{p(n-1)}\right) \approx \frac{p(n) - p(n-1)}{p(n-1)},$$

And so,

$$R_t \approx r_t$$

Hence, GR can be approximated by AR, when the market is less volatile.

To visualise the difference and to check how close AR and GR are, R_t was plotted against r_t in a scatterplot and the function $y = x$ is superimposed on the points. Points lying on the line indicate where R_t is equal to r_t (Figure 2).

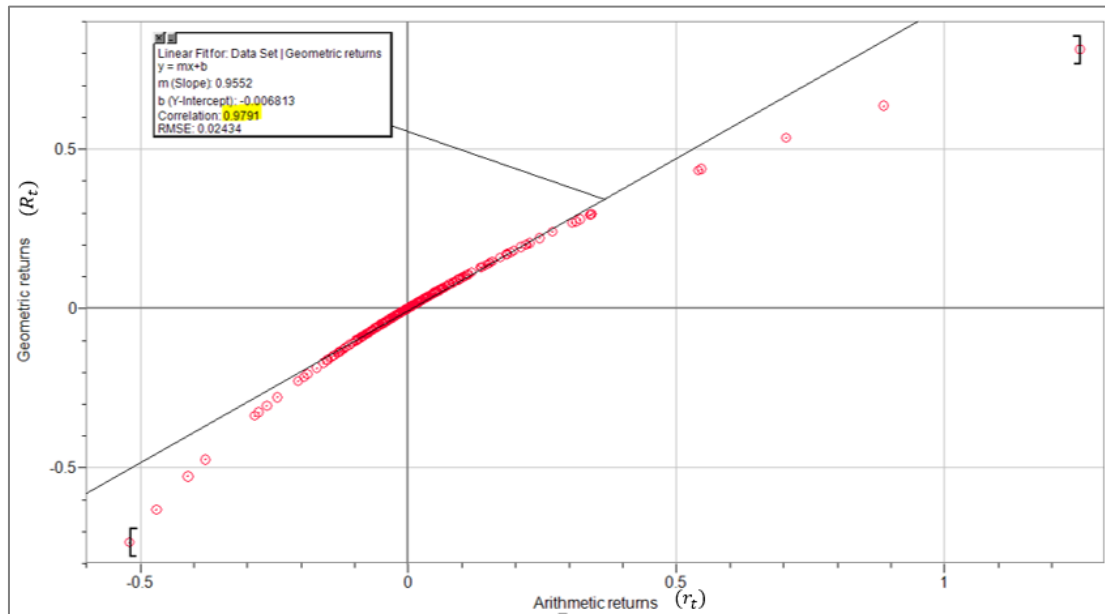


Figure 2: Geometric returns vs. Arithmetic returns

It is seen that when $|r_t| \ll 1$, the deviation from the straight line is minimal but increases as $|r_t|$ increases.

Assuming the returns follow a normal distribution, a process called standardizing is used. This sets normal variables at the same scale and returns a *z-score* for each data-value x that represents the number of standard deviations x lies from the mean. The z-score of the variable x is given by:

$$z = \frac{x - \mu}{\sigma}$$

Where, μ is the mean and σ is the standard deviation.

From the data, $\mu_{r_t} \approx \mu_{R_t} \approx 0$, $\sigma_{r_t} = 0.1387$ (4 s.f.). 95% of AR lies between $\pm 1.96\sigma$ of μ_{r_t} (Section 1.2) which is

$$\begin{aligned} &= 1.96 \times (0.1387) \text{ of } 0 \\ &= \pm 0.272 \text{ of } 0 \end{aligned}$$

$\sigma_{R_t} = 0.1268$ (4 s.f.). Hence, 95% of GR lies between

$$\begin{aligned} &\pm 1.96 \times 0.127 \text{ of } 0 \\ &= \pm 0.250 \text{ of } 0 \end{aligned}$$

As seen in Fig. 2, when $r_t \in [-0.2, 0.2]$, $r_t \cong R_t$.

$$\text{The z-score at } 0.2 \text{ for } r_t = \frac{0.2-0}{0.139} = 1.44$$

The standard normal probability at $z = 1.44$ is 92.51%

$$\text{z-score at } -0.2 \text{ for } r_t = \frac{-0.2-0}{0.139} = -1.44$$

The standard normal probability at $z = -1.44$ is 7.49%

Therefore, the probability of obtaining a value of AR in the range -0.2 to 0.2 is

$$92.51\% - 7.49\% = 85.02\%$$

Thus, 85.02% of AR are approximately equal to the GR. In general, the difference between AR and GR is negligible (Benth, 2000).

Due to the practicality of percentage values in data analysis, AR was selected to test normality.

3 NORMALITY TESTING OF ARITHMETIC RETURNS (AR)

To get an extensive set of results, normality was tested both qualitatively and quantitatively. The methods used for normality testing are given below.

1. Qualitative testing
 1. Visual evaluation of density histogram
 2. Normal Quantile-Quantile (Q-Q) Plot
2. Quantitative testing through hypothesis testing
 1. Kolmogorov-Smirnov test
 2. Anderson-Darling test
 3. Shapiro-Wilk test

3.1 QUALITATIVE TESTING

3.1.1 Visual Evaluation of Density Histogram

A density histogram is an estimation of the PDF of a continuous variable, which sorts data into equal-sized bins and the height of the vertical columns shows the density for each bin. The area under a density histogram must always be 1, so density is found by:

$$\frac{c}{nw}$$

Where c is the count for each bin, n is the total number of observations, w is interval width. Therefore, the density for a bin may be greater than 1.

Choosing a good number of bins is important to create an interpretable density histogram. The greater the number of bins, the smaller the interval, meaning that the density for each bin is high (Figure 3). The smaller the number of bins, the greater the interval, meaning that the relative frequency for each bin will be higher, but the interval will also be greater. Hence, the density will be lower (Figure 4).

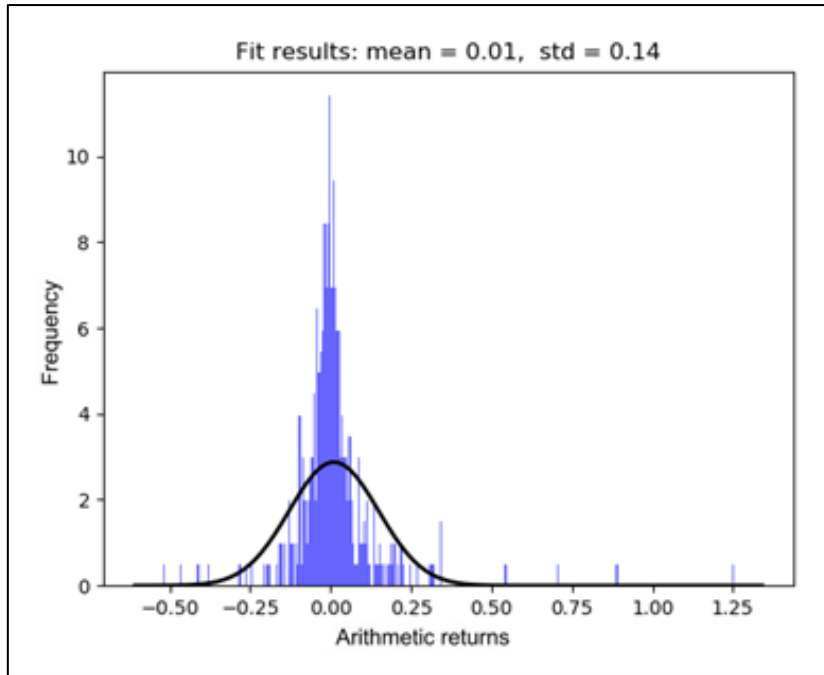


Figure 3: Histogram of AR with 320 bins

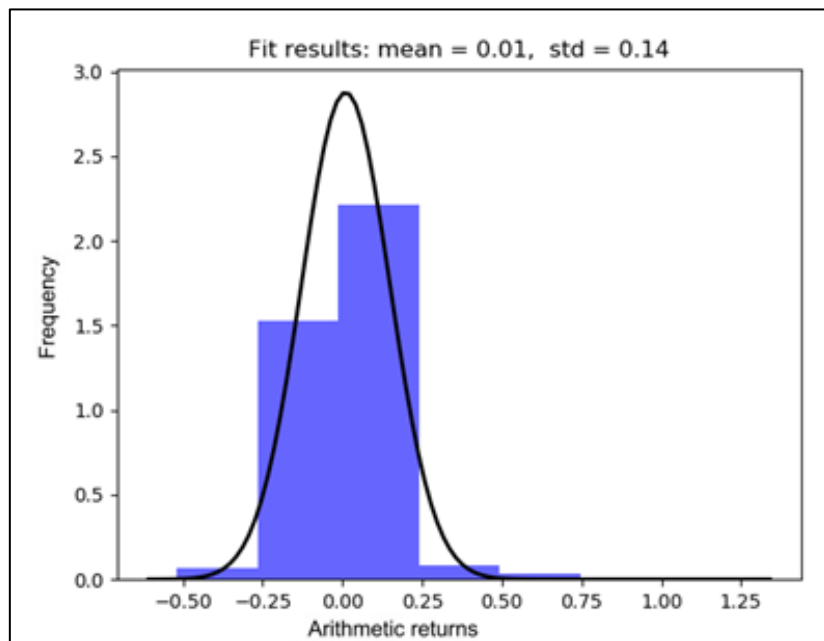


Figure 4: Histogram of AR with 7 bins

With experimentation, 40 number of bins were found to produce a good histogram. A normal PDF was superimposed over the histogram, given by the parameters found from the data:

$$P(x; 0.01, 0.14) = \frac{1}{0.01\sqrt{2\pi}} e^{-\frac{(x-0.01)^2}{2 \times 0.14^2}}$$

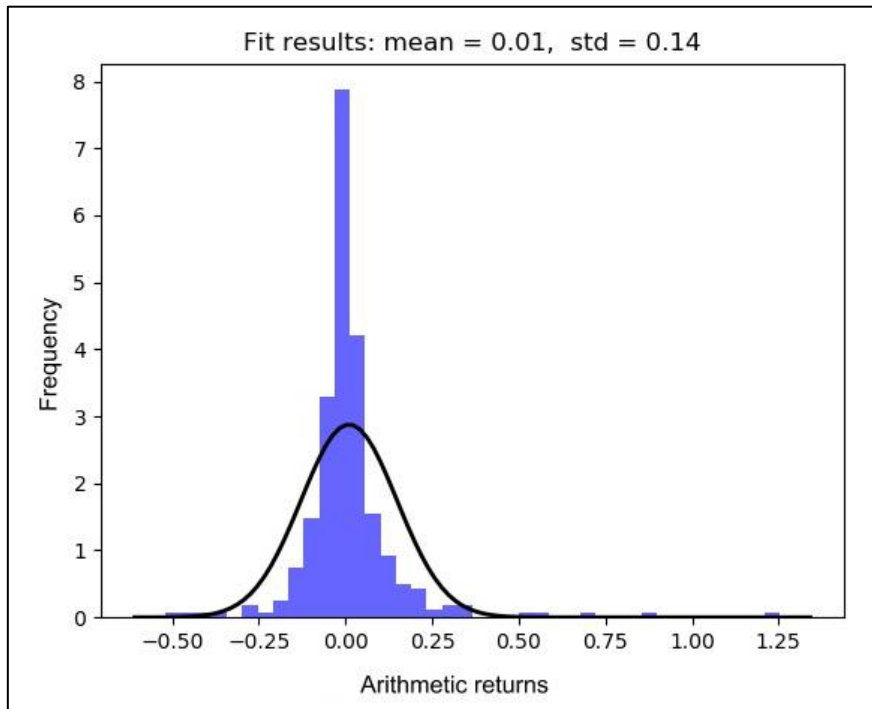


Figure 5: Histogram of Arithmetic returns with 40 bins

The histogram of AR (Fig. 5) seems unimodal, a characteristic of the normal distribution, but leptokurtic. This means that the steepness of the peak is beyond the normal PDF and is more concentrated about the mean. It seems positively skewed, which does not make sense in a power market unless the economy is expanding. Positive returns should be as frequent as negative returns in a large dataset. The skew may be explained by the presence of positive outliers.

3.1.2 The Normal Q-Q Plot

A normal Q-Q plot is a visual tool that helps to assess whether a data set comes from a normal distribution by creating a scatter plot where quantiles of the first data set are plotted against the corresponding quantile of the standard normal distribution (s.n.d.). A quantile is a value that marks where the sample covers that fraction of data. The first data set is typically the data being assessed, while the second is the z-scores occupying the same quantile. For instance, the 0.5th quantile of the data sample is plotted against the 0.5th quantile of the s.n.d. The 0.5th quantile corresponds to the median. As the s.n.d has its mean and median centred at 0, the median of the data is plotted against 0.

The steps by which a Q-Q plot is produced is shown below:

1. The observed values are arranged in increasing order.
2. The quantile that each ordered value occupies is recorded
3. From a normal distribution, find the z-scores that occupy the same quantile
4. Plot each ordered value against the corresponding z-score/theoretical quantile.

If the data sample comes from a normal distribution, the points should follow a linear trend. This is because the smallest data value is paired up with the smallest data value expected from an s.n.d of the same size and this would be done for each data-value. If the quantiles of two s.n.d are plotted against each other, it would form a 45-degree angled straight line. As any normal distribution is just a linear transformation of an s.n.d, there would only be a change in the slope of the line. However, systematic deviations from linearity reflect that the data is sampled from a non-normal distribution.

To test this, Simple Random Samples (SRS) of equal size were pulled from a normal distribution using Python and a linear regression line was passed through the data points (Figure 6).

It is seen that the Q-Q plot of two normal SRSs forms a straight line with only a few minor deviations.

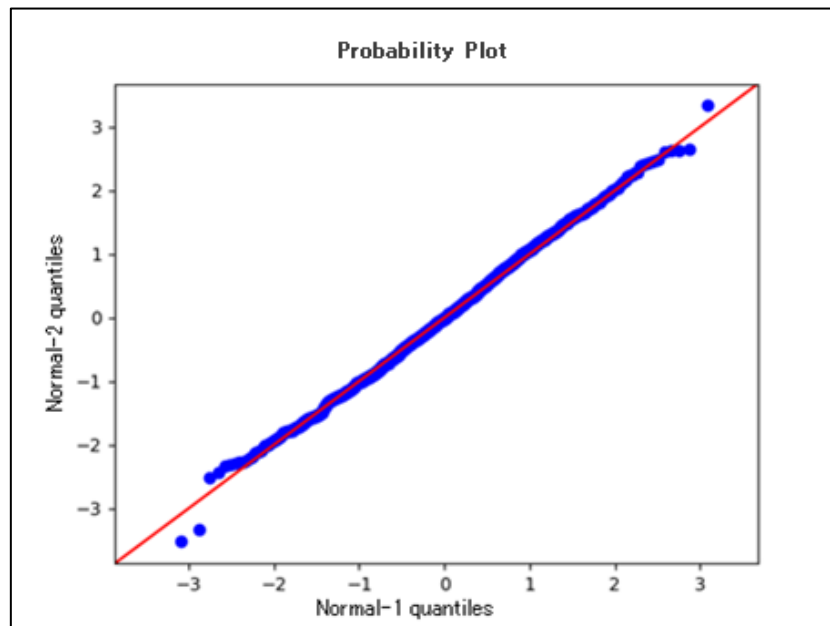


Figure 6: Q-Q Plot for two normal SRSs

A Q-Q plot for AR was then created (Figure 7).

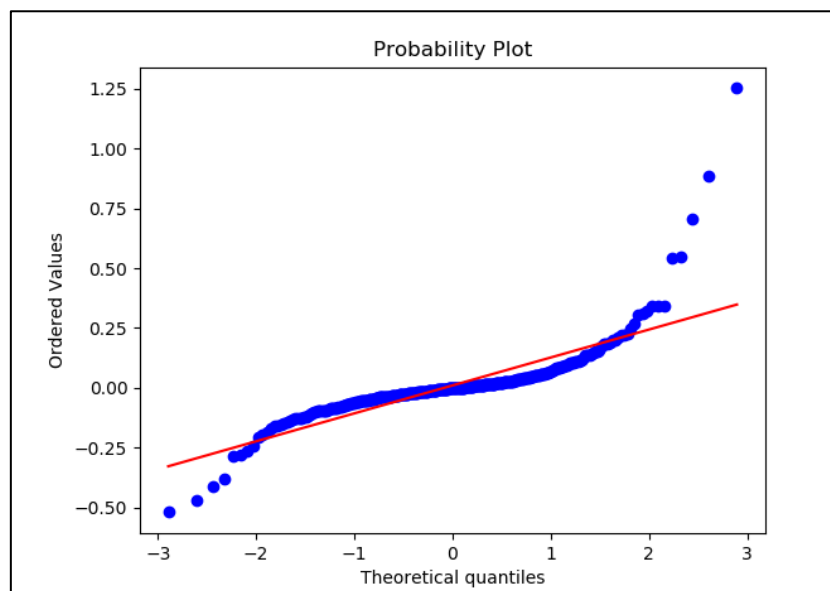


Figure 7: Q-Q plot of Arithmetic Returns

Systematic deviations from the line towards positively and negatively large quantiles can be seen. Dissecting the Q-Q plot using a guide (Figure 8) indicates that the distribution is rather heavy-tailed.

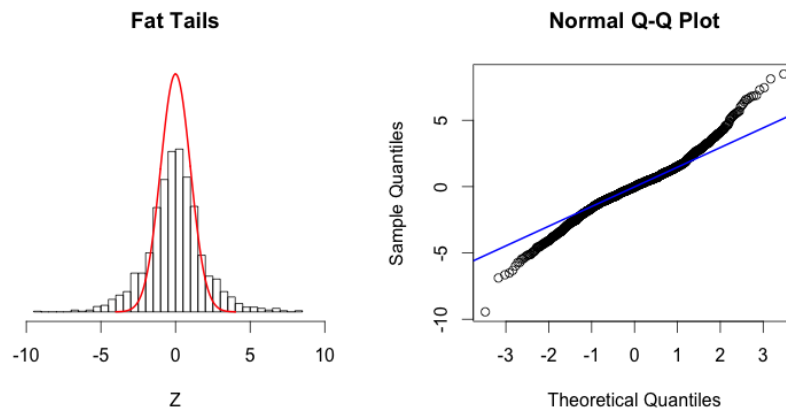


Figure 8: Q-Q discerning guide (Kross, 2016)

The guide shows a normal PDF overlaid on a heavy-tailed histogram. It can be discerned that the frequency of occurrence towards the extremes is larger than expected by a normal distribution. This means that the first sampled quantile is smaller than the first theoretical quantile and the last sampled quantile is greater than the last theoretical quantile. This leads to a systematic deviation from linearity towards the tails in the plot.

When returns are assumed to be normal when they are actually heavy-tailed, it could lead to a momentous **underestimation of risk**. This normality assumption has had massive dangers in other financial models such as the Value-at-Risk (VaR) technique. This assumption has previously made banks too sanguine, failing to forecast the 2008, U.S. housing crisis, and thus leading to negative economic spiralling (Alloway, 2012).

3.2 HYPOTHESIS TESTING

Hypothesis testing using p-values **quantify** deviations from the normal distribution and are known as goodness-of-fit tests. Before the test, the null hypothesis (H_0) and the alternative hypothesis (H_A) were decided. The null hypothesis is to be tested and aimed to be disproved. If the null hypothesis is disproved, the alternate hypothesis is accepted. A p-value provides information regarding how much evidence is in support of the null hypothesis and helps determine the significance of the results.

3.2.1 Declaration of Hypotheses

H_0 : The data-set is normally distributed

H_A : The data-set is **not** normally distributed

The p-value is typically compared against the threshold $\alpha = 0.05$.

- If $p \leq \alpha$: reject H_0 , accept H_A
- If $p \geq \alpha$: Fail to reject H_0

3.2.2 Choice of Normality Tests

There are several techniques to analyse the goodness-of-fit of a distribution of a population, given a sample. Each goodness-of-fit test returns a test statistic that is calculated differently from other tests. It is then compared against a null distribution to return a p-value, where the null distribution is test-specific. Therefore, a p-value is, assuming the null hypothesis is true, the probability that a value is found to be as extreme or more extreme than the test statistic found, by evaluating against the null distribution.

The Kolmogorov-Smirnov, Anderson-Darling and Shapiro-Wilk tests were used, which are ranked as the most powerful (Razali and Yap, 2011). These tests were chosen to achieve higher accuracy in concluding the goodness-of-fit.

A test may either be one-tailed or two-tailed. A one-tailed test only gives the probability of a deviation from the null-hypothesis, in one direction (positive or negative). A two-tailed test is concerned with both directions, therefore returning a greater p-value. The goodness-of-fit tests conducted were two-tailed, as both positive and negative deviation from normality must be considered. The tests were conducted using Python.

3.2.3 Kolmogorov-Smirnov Test (K-S)

A Kolmogorov-Smirnov test is used to compare a sample with a reference distribution, in this case, it was the normal distribution. The sample is first standardized and then its empirical cumulative distribution function (EDF) is plotted together with the normal CDF. The supremum deviation, the least upper-bound, between the two functions is quantified in terms test statistic D , which is given by:

$$D = \sup_x |f_n(x) - f(x)|$$

Where $f_n(x)$ is the EDF and $f(x)$ is the CDF.

The D test statistic is compared against the Kolmogorov null distribution to return a p-value. In Table 2, it is seen that the p-value obtained from this test was 2.731×10^{-54} . This p-value is $\ll \alpha$, and the **H₀ can be rejected**.

It is important to note, however, that the K-S test calculates its p-value only using the supremum deviation between the CDF and EDF but does not account for **where** this deviation took place, and what significance it may hold. For instance, it does not add weight to the distribution tails. As seen in the Q-Q plot, the data clearly has tails heavier than the normal distribution, meaning that a deviation in the tails of the distributions should have a higher significance. Also, the supremum deviation may be the result of an outlier that does not represent the sample data wholly.

3.2.4 Anderson Darling Test (A-D)

The Anderson-Darling test is a modified version of the K-S test and is more powerful as it adds weight to the tails of the distribution. This is especially important, seeing the heavy-tailed nature of the returns on the Q-Q plot. Deviations in tails should be weighted, as those indicate larger financial returns than that expected by a normal distribution. A feature of the A-D test is that it returns a list of critical values corresponding to the significance levels: 15%, 10%, 5%, 2.5% and 1%, along with a test statistic 'A2', rather than returning a single p-value. If the A2 is greater than the critical values, the null hypothesis can be rejected (SciPy, 2019). The test returned five, strong rejections of the H_0 (Table 2). The test statistic is much larger than all the 5 critical values.

3.2.5 Shapiro-Wilk Test (S-W)

Researchers often recommend the Shapiro-Wilk test as the best choice for normality testing (Ghasemi and Zahediasl, 2012). This test detects departures from normality through skewness and kurtosis. A weakness is that it doesn't work well on data that has identical observations. This should not be a problem for this data, as it was not found to have repeated observations.

The S-W test returned a p-value much smaller than α , thereby rejecting H_0 .

Name of Test	Test statistic (3 d.p.)	Critical values (3 d.p.)	p-value	Reject/Fail to reject H_0
Kolmogorov-Smirnov	0.404	-	2.731×10^{-54}	Reject
Anderson-Darling	25.419	1. 0.570	-	1. Reject
		2. 0.649		2. Reject
		3. 0.779		3. Reject
		4. 0.908		4. Reject
		5. 1.080		5. Reject
Shapiro-Wilk	0.712	-	1.165×10^{-24}	Reject

Table 2: Results of Hypothesis Testing, AR

All the tests clearly indicate departure from normality, and the null hypothesis can be rejected for AR.

4 NORMALITY TESTING OF GEOMETRIC RETURNS (GR)

Upon further research, it was found that the Black-Scholes formula assumes GR in its formula and not AR (Ross, 2006). Although both returns were found to be approximately equal, the formal normality-test procedure is conducted on GR to confirm non-normality and further, only GR is tested for independence and used to conclude the results.

4.1 QUALITATIVE TESTING OF DATA

4.1.1 Visual Evaluation of Histogram

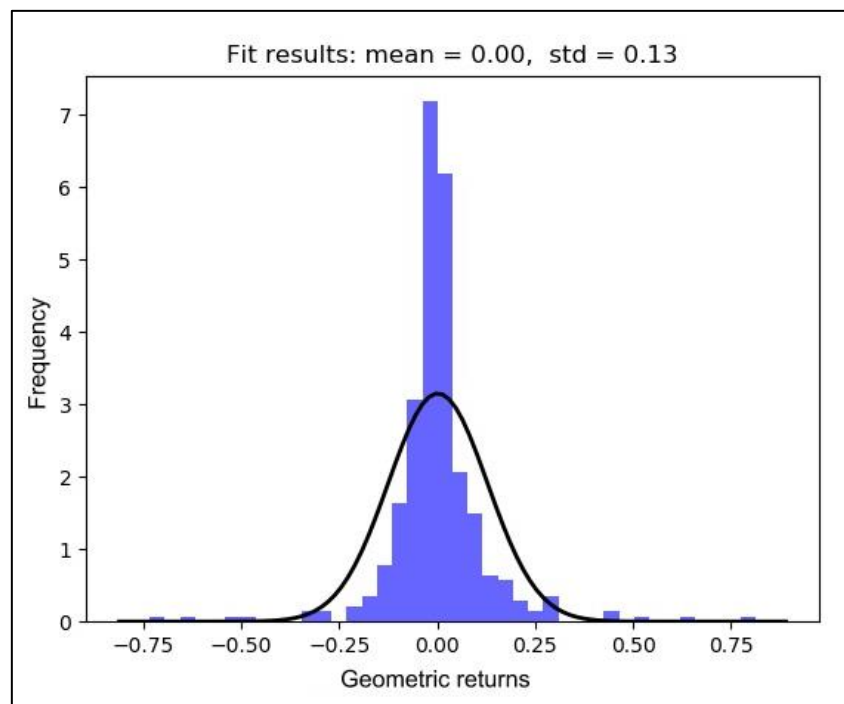


Figure 9: Histogram of geometric returns

GR also seems leptokurtic (Figure 9). However, there is no visible skew as outliers affecting the skew in AR would not have the same effect on skew in GR. This may be because GR is a logarithmic transformation of AR which reduces the scale of large values resulting in a distribution more symmetric about the mean.

4.1.2 Normal Q-Q Plot

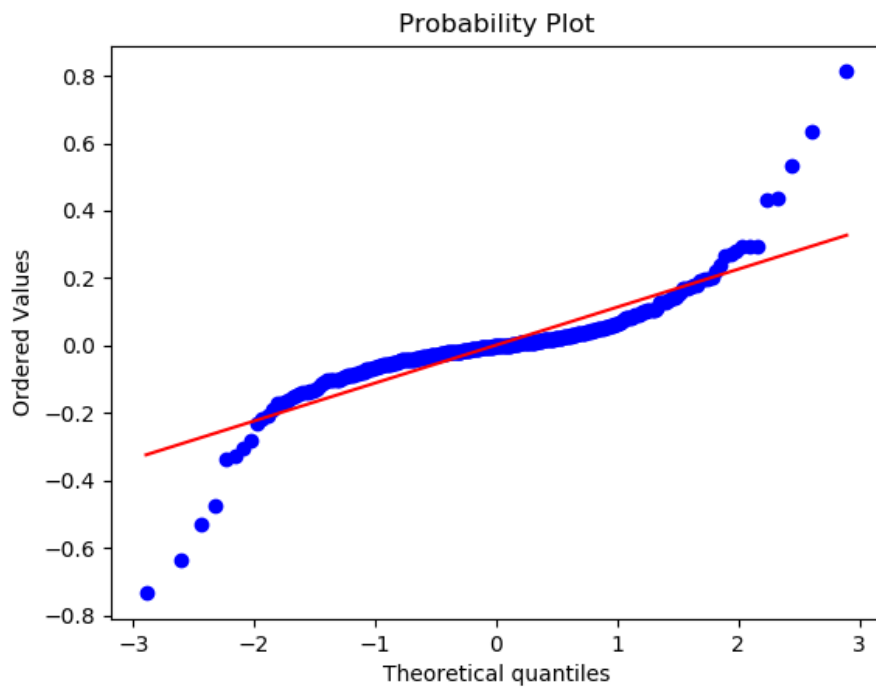


Figure 10: Q-Q plot of geometric returns

This Q-Q plot shows that GR also form a heavy-tailed distribution. The heavy-tailed nature is thus consistent in both forms of returns.

4.2 HYPOTHESIS TESTING

The Kolmogorov-Smirnov, Shapiro-Wilk and the Anderson-Darling test conducted for the results of the hypothesis testing for GR proves to give non-normal results as the p-value is much smaller than the threshold and A2 is greater than the critical values (Table 3).

Name of test	Test statistic (3 d.p.)	Critical values (3 d.p.)	p-value	Reject/Fail to reject H ₀)
Kolmogrov-Smirnov	0.399	-	8.738×10^{-53}	Reject
Anderson-Darling	20.373	1) 0.570	-	1) Reject
		2) 0.649		2) Reject
		3) 0.779		3) Reject
		4) 0.908		4) Reject
		5) 1.080		5) Reject
Shapiro-Wilk	0.788	-	0.000	Reject

Table 3: Results of Hypothesis Testing, GR

*It can, therefore, be concluded that the returns for the power market data are strongly non-normal and **do not** abide by the first assumption of the Black-Scholes formula.*

5 INDEPENDENCE OF RETURNS

A major premise behind the Black-Scholes formula is that returns follow geometric Brownian motion. Future returns should be independent of past returns, and therefore uncorrelated. If this is true, GR will be independent, identically distributed (i.i.d.) variables. To assess this, a sample autocorrelation function was created.

Correlation is a function of covariance, where covariance measures the direction of the relationship between two variables X and Y , while correlation measures its strength.

5.1 COVARIANCE

Covariance of any two variables X and Y , is given by:

$$Cov(X, Y) = E[(X - \bar{X})(Y - \bar{Y})]$$

Where,

$\bar{X} = E(X)$, the expected value of X

And,

$\bar{Y} = E(Y)$, the expected value of Y

In a positive relationship, when $X < \bar{X}$, it follows that $Y < \bar{Y}$, and the product is positive, and when $X > \bar{X}$ it follows that $Y > \bar{Y}$, and the product is again positive. Hence, the covariance of a positive relationship is positive.

In a negative relationship, when $X < \bar{X}$, it follows that $Y > \bar{Y}$ and when $X > \bar{X}$, it follows that $Y < \bar{Y}$. Thus, the product and the covariance are negative.

5.2 PEARSON'S CORRELATION COEFFICIENT

Pearson's correlation coefficient of X and Y is given by:

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

$$Cor(X, Y) = \frac{E[(X - \bar{X})(Y - \bar{Y})]}{\sigma_X \cdot \sigma_Y}$$

Correlation ranges between -1 and +1, where -1 indicates a strong negative correlation, and +1 indicates a strong positive correlation. A correlation of 0, indicates no correlation.

5.3 AUTOCORRELATION OF RETURNS

Autocorrelation measures the correlation of a series with a delayed copy of itself and can be measured at various lag lengths. Independent returns would imply that there is insignificant autocorrelation in its time series (Section 1.1.1). The measurement of significance is explored further ahead.

5.3.1 Correlation of Consecutive Returns

To measure the dependency between consecutive returns, the returns were shifted by a lag of 1, using a circular-shift function. The two variables are denoted as:

$$R_t: \text{GR}$$

$$R_{t+1}: \text{GR shifted by a lag of 1}$$

Hence, the correlation between R_t and R_{t+1} is given by:

$$\text{Cor}(R_t, R_{t+1}) = \frac{E[(R_t - \mu)(R_{t+1} - \mu)]}{\sigma_{R_t}^2}$$

As both variables are from the same series, the mean and standard deviation is the same for both.

In Figure 11, R_{t+1} is plotted against R_t and the line of regression is laid over the points.

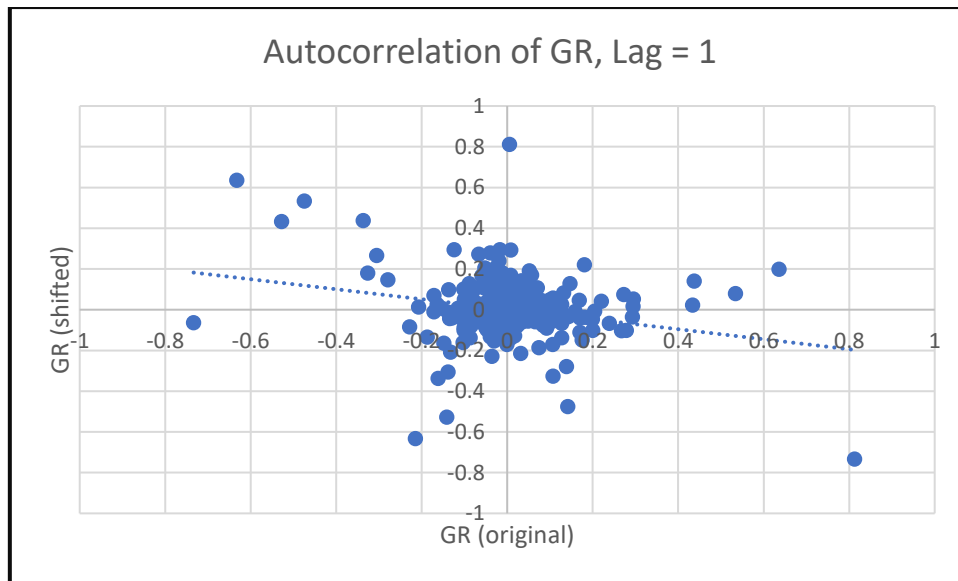


Figure 11: Autocorrelation of GR of lag 1

Visually discerned, there is a weak, but evident, correlation between R_t and R_{t+1} . Pearson's correlation coefficient was calculated to be -0.245. The presence of correlation makes sense due to the phenomenon of volatility clustering which states that large changes in price tend to cluster together (Cont, 2005), but opposes the Efficient Market Hypothesis.

To test for geometric Brownian motion, this correlation coefficient must be compared against a certain threshold-value to either reject or accept the assumption. For this purpose, a sample autocorrelation test was conducted.

5.3.2 The Sample Autocorrelation Function

The sample autocorrelation function measures the correlation between the data at increasing lags (h).

Hence, the variable shifted by lag (h) can be denoted by R_{t+h} and the autocorrelation function can be defined as:

$$Cor(R_t, R_{t+h}) = \frac{E[(R_t - \mu)(R_{t+h} - \mu)]}{\sigma_{R_t}^2}$$

The correlation was measured up to 40 lags to capture any memory in the data as financial markets may see dependencies in, for example, every 5th day which may not be evident in a lag of 1.

$$h = 0, 1, \dots, 40$$

If GR are normal i.i.d., the corresponding sample autocorrelations would also be normal i.i.d (Brockwell and Davis) with a mean 0, and variance given by:

$$s^2 = \frac{1}{n}$$

Where s is the standard deviation of the sample autocorrelations, and n is the sample size (Brockwell and Davis). The sample size of returns is 364. Hence,

$$s^2 = \frac{1}{364}$$

$$s = \frac{1}{\sqrt{364}}$$

Out of the sample autocorrelations, 95% should fall between the confidence intervals:

$$\begin{aligned} & \pm 1.96s \\ & = \pm \frac{1.96}{\sqrt{364}} \\ & = \pm 0.1027 \text{ (4 d.p.)} \end{aligned}$$

It is therefore expected that 5% of the sample autocorrelations will fall outside the bounds. As there are 40 lag-lengths, the number of sample autocorrelations expected to fall outside the interval, if returns are normal i.i.d. is,

$$= 40 \times 0.05$$

$$= 2$$

If any more observations fall outside the bounds, this would be statistically significant and the premise can be rejected. Figure 12 shows the sample autocorrelation function generated using MATLAB where the red marks represent the lag lengths at which the autocorrelation is beyond the bounds.

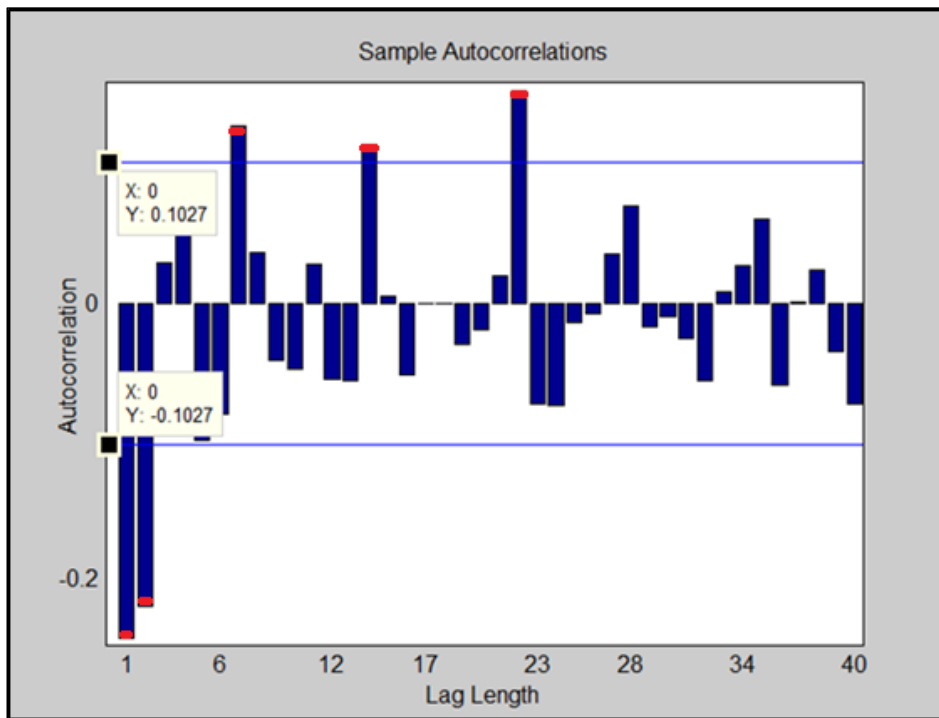


Figure 12: Sample autocorrelations for GR up to 40 lags

Figure 12 shows 5 sample autocorrelations falling outside the bonds, marked in red. As this number is greater than 2, it is a statistically significant number and it is possible to conclude that GR have a sufficiently strong correlation. However, correlation does not imply causation, and some correlation may be coincidental. It can therefore only be concluded with **95% confidence** that *the returns are not normal i.i.d. and consequently, do not follow geometric Brownian motion. The second assumption of the Black-Scholes formula is thus, not satisfied.*

6 CONCLUSION

This report has established that the Oslo, 2018 power prices do **not** satisfy the assumptions of normality and independence of the widely-used Black-Scholes option-pricing formula. Through qualitative and quantitative tests, it was found that both arithmetic and geometric returns are significantly non-normal and heavy-tailed and will continually cause an underestimation of risk in financial modelling. The power prices were shown to not follow geometric Brownian motion by applying an autocorrelation test of 95% confidence.

The year, 2018, could be a limitation of the test as it may be abnormal in terms of its behaviour, however, further information regarding this could not be found. Perhaps by the central-limit theorem, increasing the sample size to multiple years would produce a sample better representative of the population of power prices in Norway. However, doing this could introduce seasonality which would require treatment. Additional normality tests can also be conducted to confirm the results which strongly reject normality.

The findings are significant as they show that the power market does not adhere to the assumptions of the Black-Scholes formula, meaning that the option price derived from the formula will not be risk-neutral or eliminated of the opportunity of arbitrage. Although both assumptions are not met, the Black-Scholes formula may still be used, but it might not fulfil the initial requirements of risk-neutral valuation. With further investigation, more samples from Norway and other countries can be tested to see if they cohere with this conclusion. Additionally, explore other option-pricing models that look beyond the assumptions of the Black-Scholes formula and accommodate the heavy-tailed nature of returns.

7 REFERENCES

1. Alloway, T. (2012). Modelling: Normal distribution is not always the norm | Financial Times. [online] Ft.com. Available at:
<https://www.ft.com/content/67d05d30-7e88-11e1-b7e7-00144feab49a>
[Accessed 13 Oct. 2019].
2. Cont, R. (2005). Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models. [online] ResearchGate. Available at:
https://www.researchgate.net/publication/228310932_Volatility_Clustering_in_Financial_Markets_Empirical_Facts_and_Agent-Based_Models [Accessed 14 Oct. 2019].
3. Ghasemi, A. and Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. International Journal of Endocrinology and Metabolism, 10(2) [online], pp.486-489. Available at:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693611/> [Accessed 20 Sep. 2019]
4. Introduction to the Black-Scholes formula. (2013). [video] Directed by S. Khan. Khan Academy. Available at: <https://www.youtube.com/watch?v=pr-u4LCFYEY>
5. Kross, S. (2016). A Q-Q Plot Dissection Kit . [online] Available at:
<https://seankross.com/2016/02/29/A-Q-Q-Plot-Dissection-Kit.html> [Accessed 7 Jun. 2016].
6. Nau, R. (n.d). Stationarity and differencing of time series data. [online] Available at: <https://people.duke.edu/~rnau/411diff.htm> [Accessed 18 Oct. 2019].
7. Razali, N. and Yap, B. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. [online] ResearchGate. Available at:
https://www.researchgate.net/publication/267205556_Power_Comparisons_of_Shapiro-Wilk_Kolmogorov-Smirnov_Lilliefors_and_Anderson-Darling_Tests
[Accessed 19 Sep. 2019].
8. Simonsen, I. (2004). Volatility of Power Markets. [online] Web.phys.ntnu.no. Available at:
http://web.phys.ntnu.no/~ingves/Downloads/MyPapers/PhysicaA_355_10_2005.pdf [Accessed 5 Jun. 2019].

Books:

- 1) Benth, F. (2000). Option Theory with Stochastic Analysis. Berlin, Heidelberg: Springer Berlin Heidelberg.
- 2) Brockwell, P. and Davis, R. (2002). Introduction to time series and forecasting. 2nd ed. New York: Springer.
- 3) Ross, S. (2006). An Elementary Introduction to Mathematical Finance. 2nd ed. Cambridge: Cambridge University Press, pp.213-239.

Methods:

1. Docs.scipy.org. (2019). scipy.stats.anderson — SciPy v1.3.1 Reference Guide.
[online] Available at:
<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.anderson.html> [Accessed 9 Oct. 2019].
2. MATLAB. (2018b).
3. Python (3.6.8)

8 APPENDIX

- A. Spot-price power data for Oslo, 2018 is shown on the next page and is taken from Nordpool historical market data:

URL: <https://www.nordpoolgroup.com/historical-market-data/>

