Investigation into the feasibility of predicting movements in the stock market from daily Covid-19 cases

James E. McDonald

January 9, 2023

Abstract

This project aims to determine whether it would have been possible to use the daily reported cases during the Covid-19 pandemic to predict the movement of share prices, and therefore profit from the stock market. To establish whether one could predict the share movements, I will test the extent to which the performance of shares in different sectors, which were especially affected by Covid-19 according to the Office for National Statistics, is correlated with the daily cases reported by the World Health Organisation. Throughout my investigation, I will also compare the results from the two methods used to find the correlation: the Interpolated Cross-correlated Function and the Discrete Cross-correlation Function, which are both used frequently in astronomy.

Contents

1 Introduction

This project aims to find a correlation between daily Covid-19 (Covid) cases reported by the World Health Organisation [4] and the price of shares to determine whether it is feasible for the data to be used to generate profit.

The stock market is a very complex system which is constantly adapting to new data, with data affecting individual share prices differently. It is this concept which the Efficient Market Hypothesis is based. It states that the stock market reflects all available relevant data therefore not making it possible for investors to make significant gains from "timing the market" [3]. However, the Covid pandemic was such a unique event, with regards to scale and in terms of data being released on a daily basis, it's unlikely that it's affects could have been immediately factored into the share prices. As a result it is reasonable to expect opportunities to profit to arise during the pandemic, as illustrated by some hedge fund managers making the biggest returns of the decade [2] despite many industries struggling.

As the Covid pandemic affected different industries uniquely, I have used share prices from Yahoo Finance [1] to compare companies in the four most affected sectors [7], transportation and storage (Easyjet), wholesale and retail (Tesco), accommodation and food services (Whitbread) and arts, entertainment and recreation (Cineworld).

Two types of cross-correlation function have been used To compare the two sets of data: the Interpolated Cross-Correlation Function and the Discrete Cross-Correlation Function. Although, these two methods for finding correlations are very well documented [8] and should theoretically provide the same results, this project includes a description of methodology used as well as explores reasons why the functions don't always generate the same answer.

2 Method

2.1 Cross-Correlation Functions

Cross-correlation functions (CCFs) provide a statistical method of comparing how well two sets of data mirror each other's movements. In this project, I will be using two types of CCF: the Interpolated CCF (ICCF) and the Discrete CCF (DCCF). Both methods use the same equation but differ in the way they apply it.

The CCFs generate values between 1 and -1 across different offsets for the two data sets. A value of 1 is the strongest correlation, which shows they perfectly match one another, while -1 shows that they are perfectly negatively correlated, meaning the data sets move in opposite directions: as one value goes up the other goes down and vice versa. A value of 0 indicates there is no correlation.

Once the correlation value has been generated for all the desired offsets, the highest and lowest value of the correlation coefficient can be found and the corresponding offsets are taken at the points where the two data sets most closely follow, or mirror, one another. If the correlation is not a coincidence then it can be used to predict the scale and direction of one data set's movement based on the other data set and the calculated offset value.

2.2 Interpolated cross-correlation function

The ICCF uses, as the name would suggest, interpolation to fill in missing data points. For example, there is no stock market data for weekend's and bank holidays as no trading takes place on these days, which reduces the amount of available data significantly.

To interpolate the data, I used the following equation:

$$y = y_1 + (x - x_1)\frac{y_2 - y_1}{x_2 - x_2} \tag{1}$$

where x and y are the interpolated values, x_1 and y_1 are the last data points before the gap in the data, and x_2 and y_2 are the first data points after the gap.

Equation 1 simply plots the interpolated data points in a straight line between the last point before the gap and the first point after the gap, but this wouldn't give a very accurate interpolation as the stock market naturally has some variation and noise within trends. To account for this, I added some simulated noise, by taking the local standard deviation over 10 days, generating a random number between 1 and -1 with a normal probability distribution and a standard deviation of 1, multiplying the two to get "noise" and adding this to the value generated by the equation above (1). I use the normal probability distribution for choosing a random number as naturally the point is more likely to be closer to the trend line than be displaced by one standard deviation, so it is reasonable to use a normal distribution when deciding how much noise is added to the point.

Once the data has been interpolated, the equation for the ICCF is used:

$$CC(\tau) = \frac{1}{N-1} \sum^{N} \frac{(x(t) - \bar{x})(y(t-\tau) - \bar{y})}{\sigma_x \sigma_y} \tag{2}$$

CC is the correlation coefficient for each offset, the offset is represented by τ and shifts one of the data sets by changing each data point's t value, which is the date corresponding to each data point. $\bar{x}$ and $\bar{y}$ are the average of each of the data sets, σ_x and σ_y are the standard deviations of each of the data sets and N is the total number of data points for which the correlation is being calculated.

Equation 2 indicates the type of correlation by calculating the difference between the mean of the data set and a point in the data set for both data sets and multiplying them together. If both values are above or below the mean, a positive correlation is generated as the product will be positive. Conversely, a negative correlation (anti-correlated) exists if one value is above and the other is below the mean while there is no correlation if one value is exactly on the mean. This product is then divided by the standard deviations for normalisation purposes. Once a value has been calculated for one specific data point in each data set, it is repeated for the remaining data points, summed over and divided by the number of points minus one to get the average correlation coefficient for a specific offset τ. τ is increased by 1 and the process is repeated until a distribution of correlation values has been created for all the desired offsets.

2.3 Discrete cross-correlation function

The DCCF works in a similar way to the ICCF except that, it doesn't interpolate over the data, so gaps remain. Further, although the DCCF uses the same equation as the ICCF, it doesn't compare each point to the corresponding shifted point; rather, it compares the first point of data set 1 to all points in data set 2 before moving onto the second point of data set 1. This results in the following equation:

$$CC = \frac{(x_i - \bar{x})(y_j - \bar{y})}{\sigma_x \sigma_y} \tag{3}$$

Each value of the CC is binned into the corresponding offset but, unlike the ICCF, this will lead to multiple values for each offset. To calculate a single value for each offset one can take the mean of each bin and use the error on mean to determine the error on each offset. This ensures the smaller offsets that will have more values (because less data is lost to being offset) have a smaller error than those with large offsets (which will have less data as more data won't have a corresponding point).

2.4 Finding errors

As mentioned previously, a large amount of noise exists within the movement of the stock market and Covid data. In order to account for this, a simulation can be used to remove a random 30 percent of the data. For the ICCF, the data is interpolated but for the DCCF the data is left with the missing data. As a result, the DCCF will fewer data points than the ICCF which causes slight differences in the results which will be discussed later in section 3.5.

Deleting the data is done a number of times each time producing a different correlation function and therefore a different maximum and minimum for the correlation values, it is equivalent to repeating an experiment in the lab to find errors. To find the overall max or min value, the average of the individual values can be taken and the error on the mean can be taken as the overall error. The method for finding these errors has been taken from the paper, "On Uncertainties in Cross-Correlation Lags and the Reality of Wavelength-dependent Continuum Lags in Active Galactic Nuclei" [6].

3 Results

3.1 Auto-correlation

An auto-correlation can be calculated to ensure that the ICCF and the DCCF are working as expected. To calculate an auto-correlation, the same data set is used for both data sets x and y instead of inserting two different data sets into the relevant CCF. This is useful as the result should be known: if it's the same data, it should have its best correlation when the offset equals zero and the correlation coefficient should equal 1 as the data is identical (assuming no noise in the interpolation and no random set of data has been deleted).

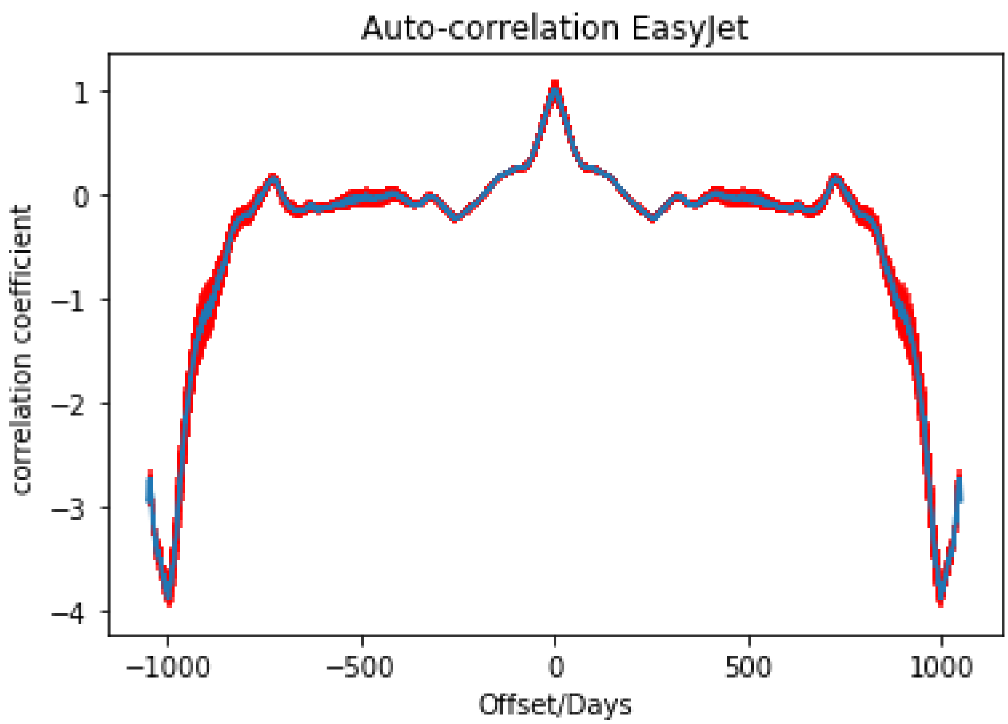

Figure 1: Auto-correlation of Easyjet's share price using the DCCF

Figure 1 (3.1) and figure 2 (3.1) show unexpected results as the correlation coefficient falls below -1 at high offsets when, as stated previously, it should remain within the range 1 and -1.

This is due to the fact that I use the mean of the whole data set instead of a local data set when calculating the ICCF and DCCF. Consequently, if there is an underlying trend in the data, the data points used when calculating high offsets are very far from the mean, resulting in very large correlation coefficients being calculated.

As it is unlikely that there will be any correlation at such large shifts, it is acceptable to ignore these large offsets and to continue using the mean for the whole data set.

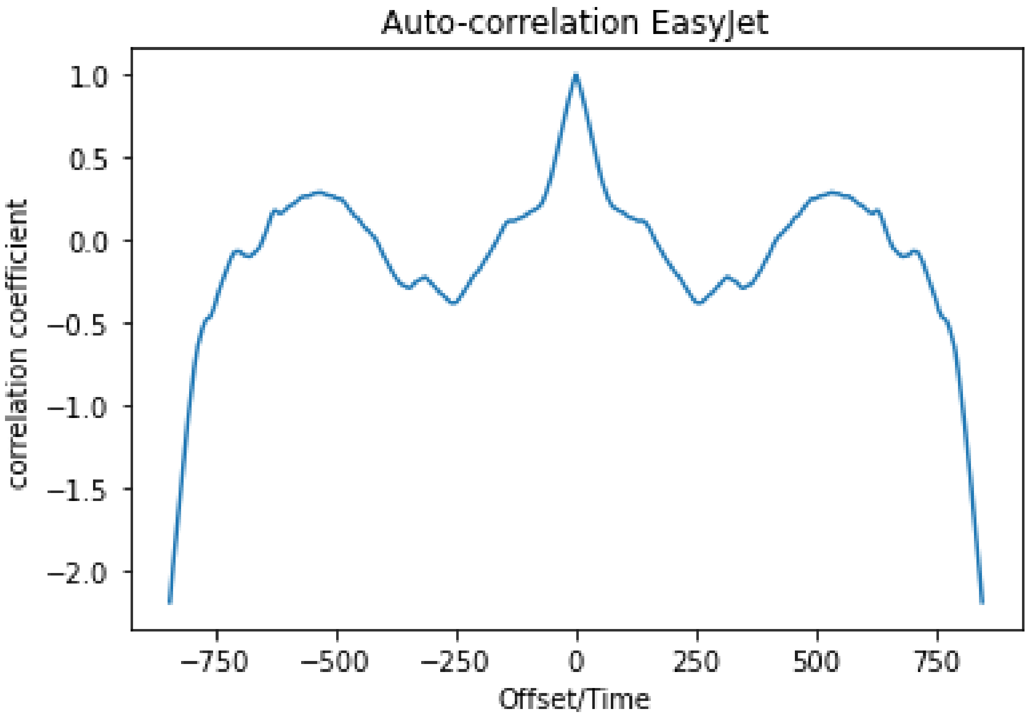

Figure 2: Auto-correlation of Easyjet's share price using the ICCF

As expected, figure 3 (3.1) and figure 4 (3.1) show that the offset peaks at 0 with a value of 1 as the ICCF has no noise in the interpolation, otherwise the two data sets wouldn't be identical and the correlation coefficient wouldn't peak at 1. This is reassuring as it means that both functions are working as expected. However, they are not identical; they have the same features but the size of these features is different. This is something I will discuss later.
It is important to note the width of the peak of the auto-correlation: the wider the peak the more 'memory' the data is said to have and the longer the trend is likely to last. Imagining a data set which has one sharp peak and is flat everywhere else, it can be easily realised how the auto-correlation of this data would have one sharp peak where it's correlated at 0 and some smaller number everywhere else. Conversely, data with a wide normal distribution-like curve will have a wider auto-correlation peak as the trend extends over a longer period.

Looking at the auto-correlation function for Easyjet in figure 4 (3.1) the peak is quite sharp and widens at the base, which indicates a very erratic nature with an underlying trend. This is confirmed in figure 5 (3.1)

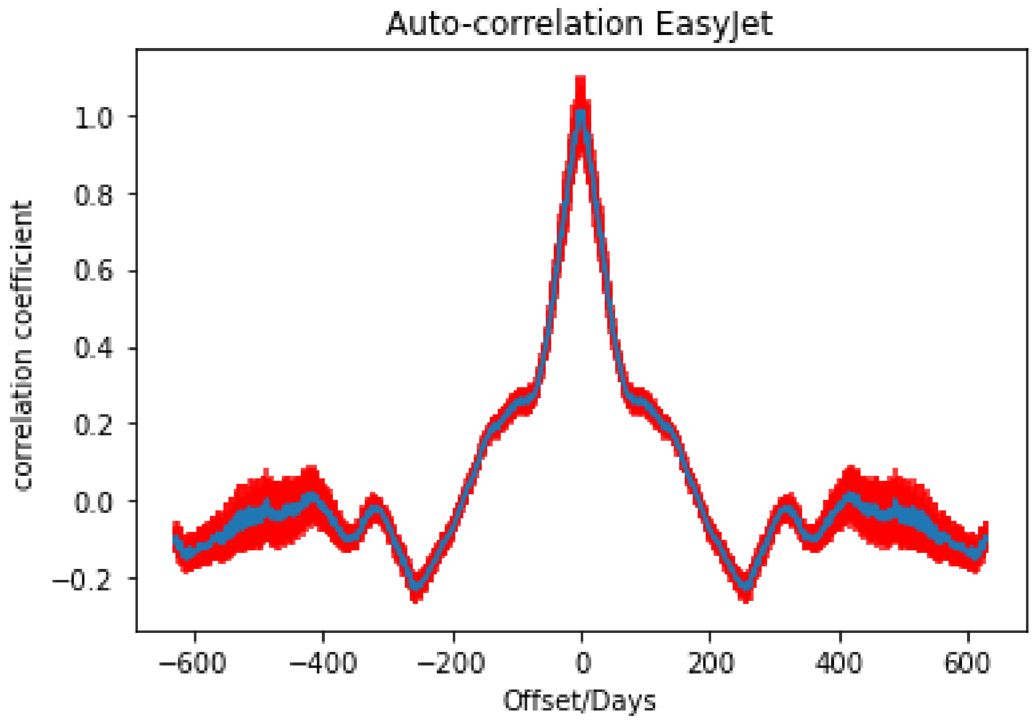

Figure 3: Auto-correlation of Easyjet's share price using the DCCF ignoring the irrelevant offsets

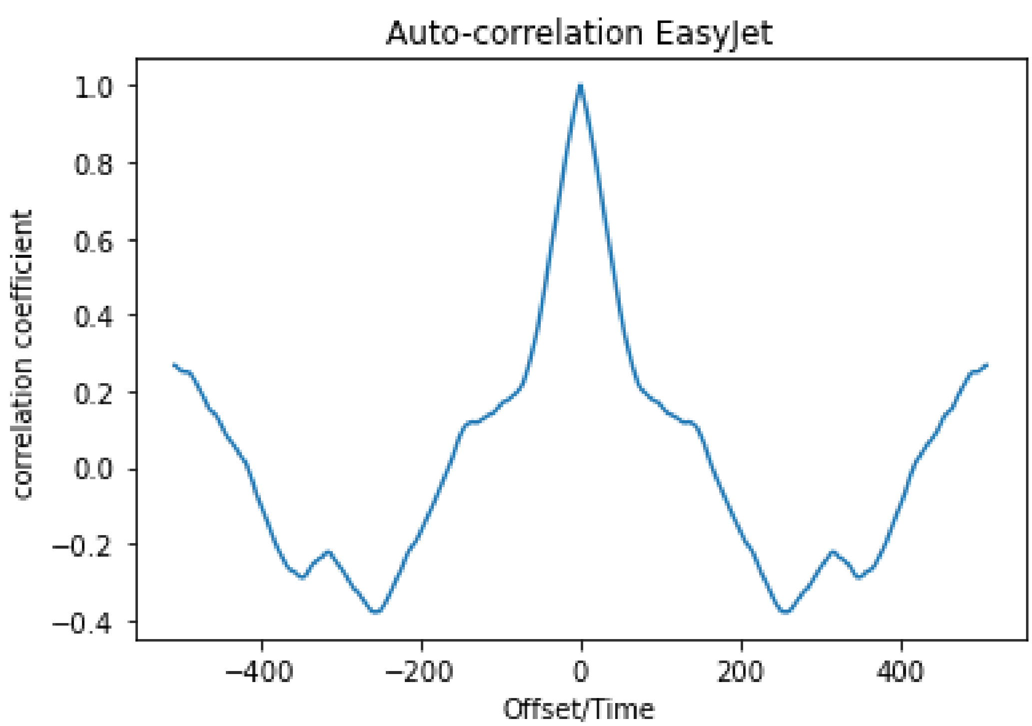

Figure 4: Auto-correlation of Easyjet's share price using the ICCF ignoring the irrelevant offsets

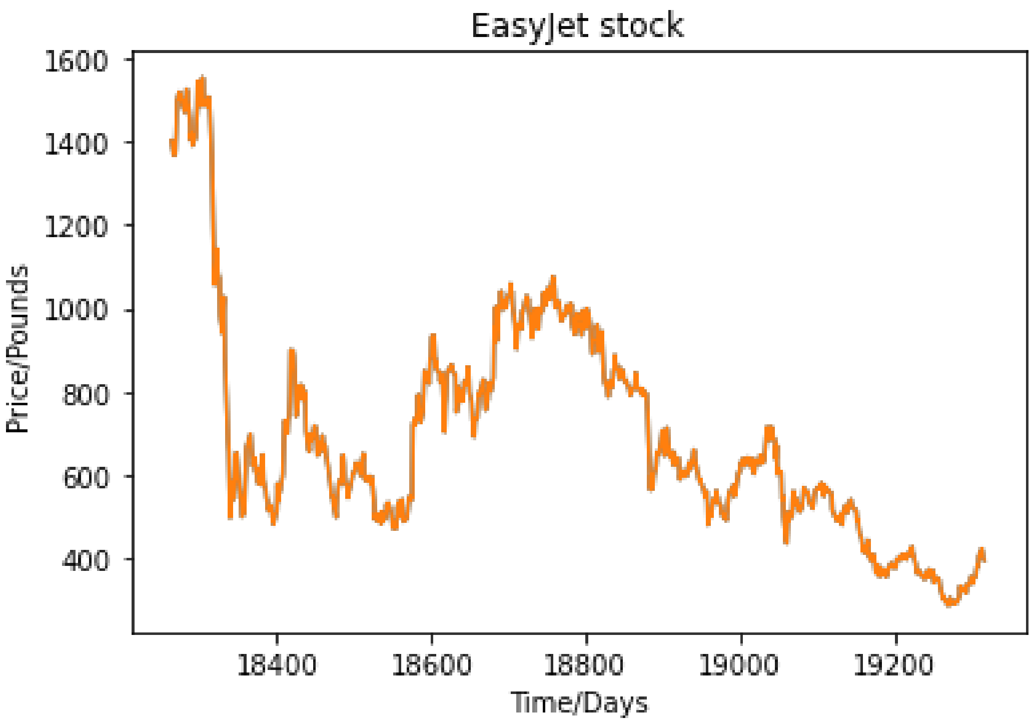

Figure 5: Time series of Easyjet's share price between 3/1/2020 and 18/11/2022

3.2 Estimating errors with simulation

Simulations are used to factor out noise in the data and get a value for the error, as illustrated in figure 6 (3.2). The graph shows the DCCF auto-correlation for Easyjet as seen previously, but with the code being run over 100 times and with 30 percent of the data points being deleted randomly each time. The code plots the DCCF each time to show the amount of deviation but the resulting graph appears quite chaotic. This is due to the fact that the DCCF doesn't interpolate data so, along with the stock market not trading on weekends, there is a limited amount of data, with even less at larger offsets. Consequently, little fluctuations in the trends have a large difference. The equivalent ICCF graph is much less spread out due to the interpolation of the data.

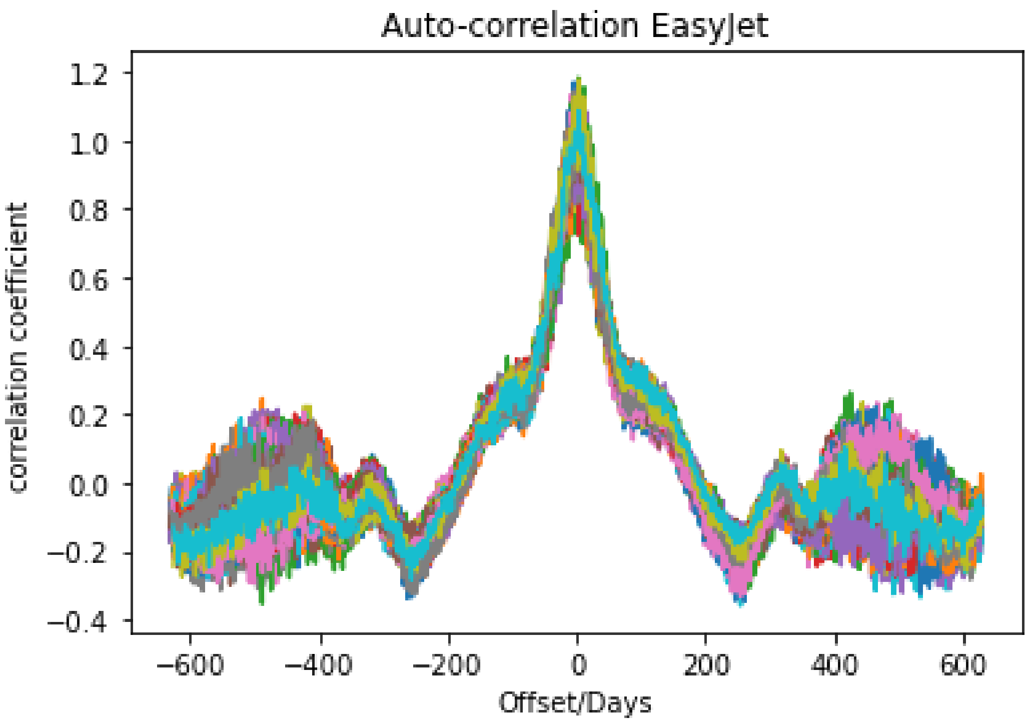

Figure 6: DCCF auto-correlation of Easyjet's share price when plotted after each simulation

When using two different data sets, most peaks are less defined and noisier than those in the auto-correlation, making it harder to find the corresponding offset indicated by the peak. To resolve this, a centroid is taken to estimate the best offset of the individual ICCFs and DCCFs. The centroid involves finding the peak y-axis value and the x-axis values either side of it that correspond to 70 percent of the peak y-axis value. With the two x-axis values either side of the peak, the average is taken to find the centroid peak.

The peak offset and correlation coefficient is then found for each of the separate simulations, the mean taken and the error on the mean used for the error. The probability distribution can then be plotted to determine the distribution of the peak offsets, as seen in figure 7 (3.2). As there is no analytical way of identifying a significant peak without using advanced simulation, the probability distribution is the best indication of how significant the peak is. If the distribution is very narrow then it can be said to be a particularly significant correlation with little error.

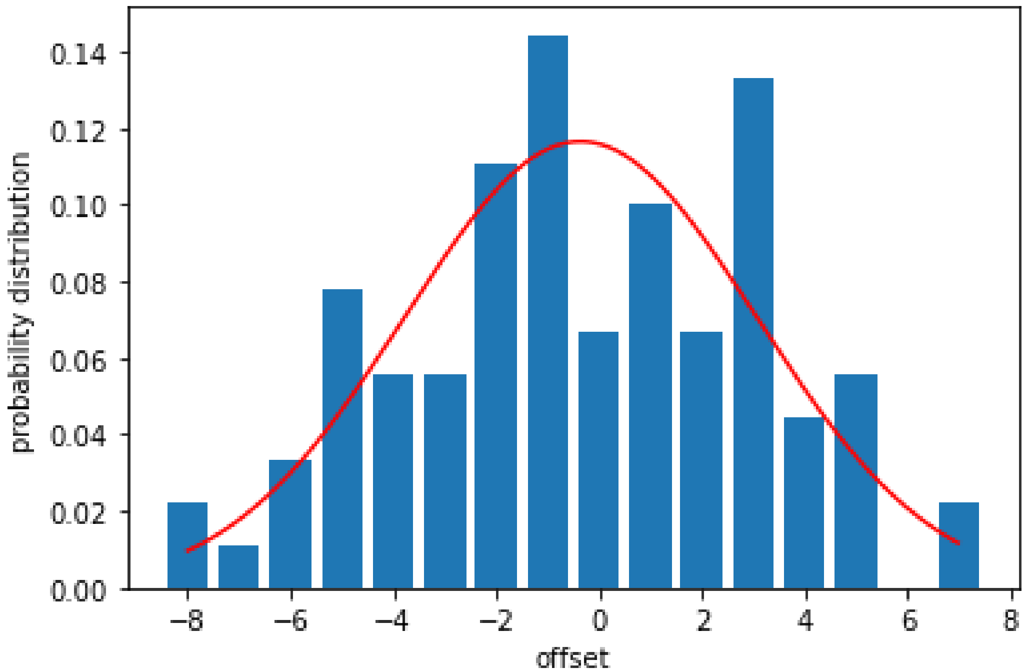

Figure 7: Probability distribution of the auto-correlation of Easyjet using the DCCF, peak offset $= -0.36 \pm 0.36$ and peak correlation coefficient $= 0.989 \pm 0.007$

3.3 Comparing Uk and European Covid-19 cases

As this report aims to find correlations between the number of reported Covid cases and movements in the price of shares, it seems logical to first see if a correlation exists between Covid cases in different regions.
It can be seen from figure 8 (3.3) and figure 9 (3.3) the DCCF has a much wider distribution than the ICCF, which is a documented characteristic of the DCCF [8]. This is caused by the fact that the DCCF doesn't interpolate over the missing data and, with 30 percent of the data being deleted and the Covid cases having quite narrow peaks, the peaks become harder to identify when points have been deleted. However, both indicate that the best offset is around 21 days, indicating that UK Covid cases are approximately 21 days ahead of the Euro cases.

Looking at figure 24 (A), it is clear that cases are quite well correlated with the shift although it has been dominated by the large spike.

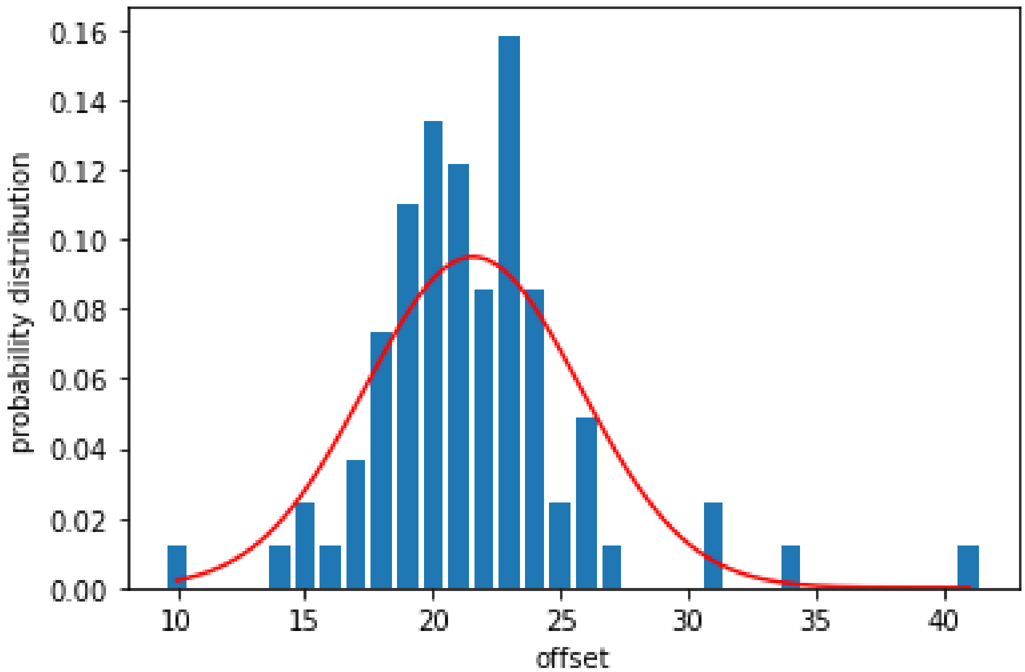

Figure 8: Distribution of offsets for the correlation between UK and Euro Covid cases using the DCCF, peak lag of Euro cases $= 21.57 \pm 0.46$ and peak correlation coefficient $= 0.83 \pm 0.009$

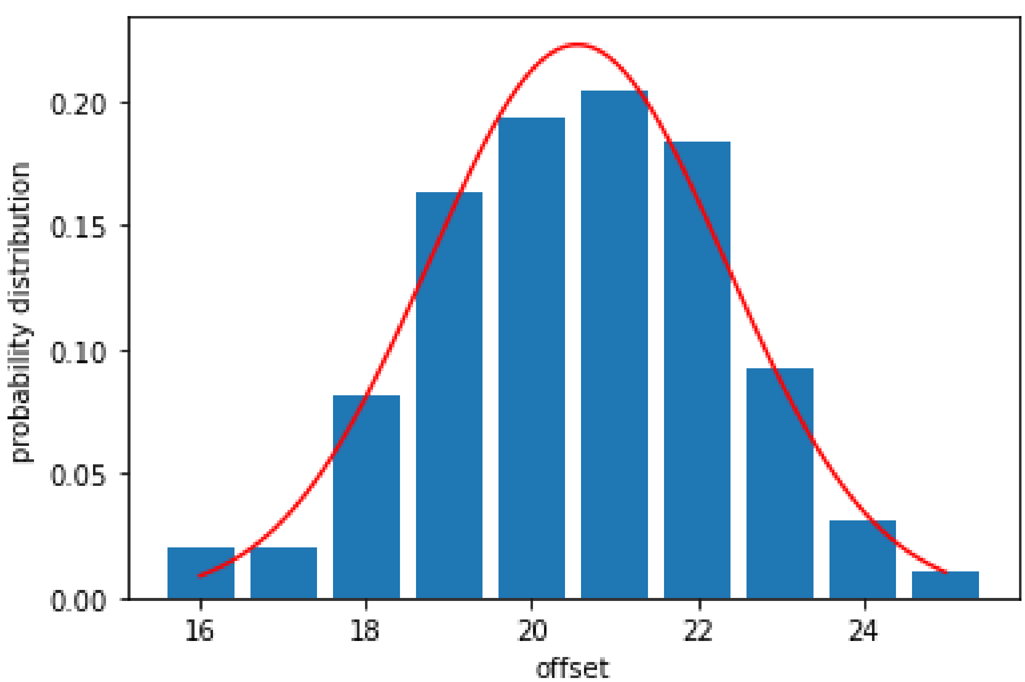

Figure 9: Distribution of offsets for the correlation between UK and Euro Covid cases using the ICCF, peak offset $= 20.55 \pm 0.18$ and peak correlation coefficient $= 0.83 \pm 0.002$

3.4 Comparing Covid-19 cases with share prices

3.4.1 Easyjet

According to the Office for National Statistics [7] there was a 78 percent decrease in turnover in air transport in the first half of 2020, which should have had a dramatic effect on airlines' share price. This hypothesis appears to be valid based on the graph in figure 10 (3.4.1).

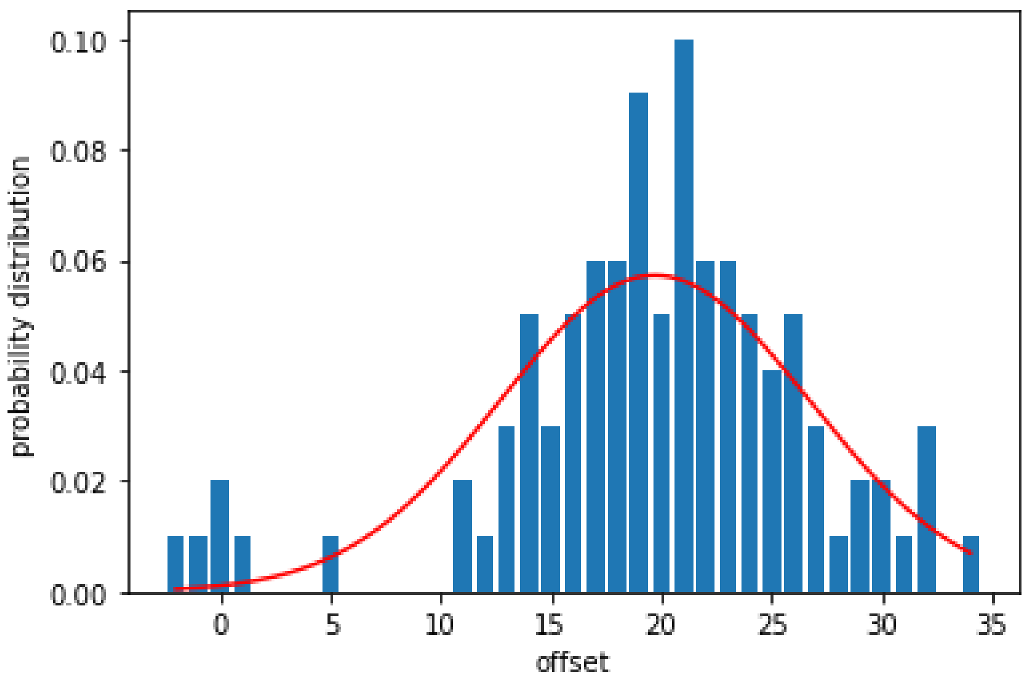

Figure 10: Probability distribution of UK cases against Easyjet's share price using ICCF, peak lag of Easyjet $= 19.7 \pm 0.7$ and peak correlation coefficient $= -0.339 \pm 0.002$

Although the correlation coefficient is low, the distribution looks well-formed, which implies that the peak of the ICCF is well defined and indicating a sufficient correlation rather than noise. To further validate the hypothesis, Easyjet share price can be plotted against UK cases.

Figure 11 (3.4.1) shows that the correlation looks correct with Easyjet's share price decreasing at the sign of an increase in Covid cases. As the correlation is best when Easyjet's share price is shifted forward by 20 days, the graph suggests that the share price was lagged behind the Covid cases by 20 days, implying one would be able to predict its movement from the daily cases.

However, testing the DCCF gives an unexpected result as it calculates the best offset to be 269.7 ± 0.8, which is nowhere near the result of the ICCF. Looking at the shifted plot of the cases against the share price with the shift gives an indication as to why this may have occurred.

Figure 12 3.4.1 shows the best correlation occurs at the shift where the lowest point and highest point are aligned. This is likely to be a result of such a large shift causing a number of data points to be excluded. Additional data points will also have been removed by the 30 percent sampling and as a result of the stock market closing on weekends and public holidays. Consequently, there will be far fewer data points when calculating the correlation function, meaning that a point with high correlation will be weighted much more heavily leading to the highest and lowest point being correlated with each other.

Ignoring the DCCF values that appear to be affected by the two large points produces an offset of 3.9, which is much closer to the ICCF although not exactly identical. As a likely result of weighting on the peak of Covid cases and the significant trough of the Easyjet share price, this offset starts to become affected by the two points being moved closer, which over-shadows the true best correlation. From viewing the probability distribution can be seen in figure 25 (A), it can be seen that the distribution is not particularly uniform, implying that it's not a significant peak and the ICCF is much more likely to be correct.

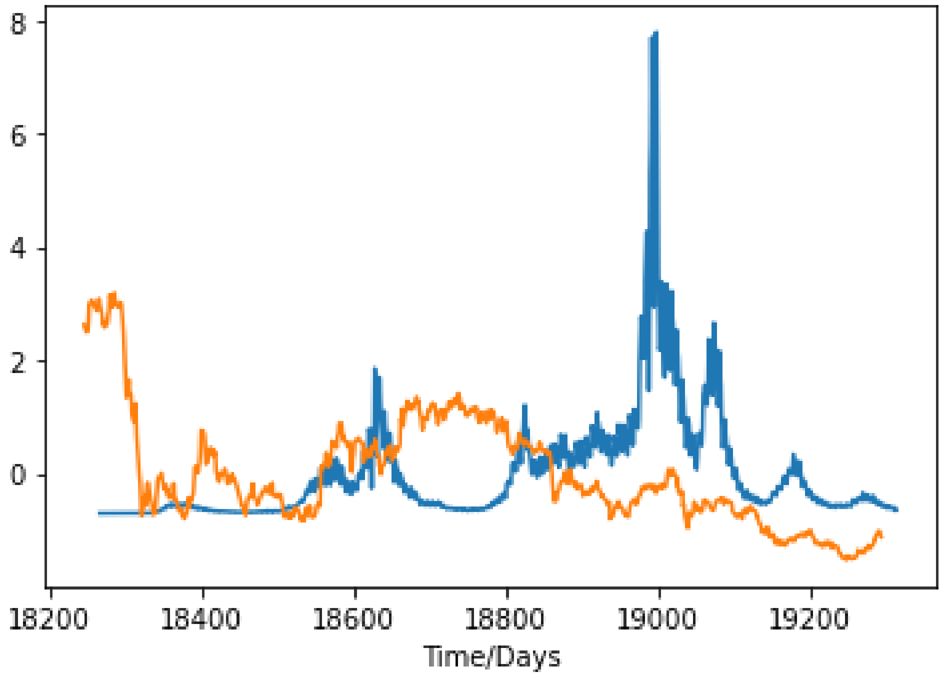

Figure 11: UK cases with Easyjet's share price (with mean taken away and divided by standard deviation for normalisation) plotted with the 20 day offset suggested by ICCF

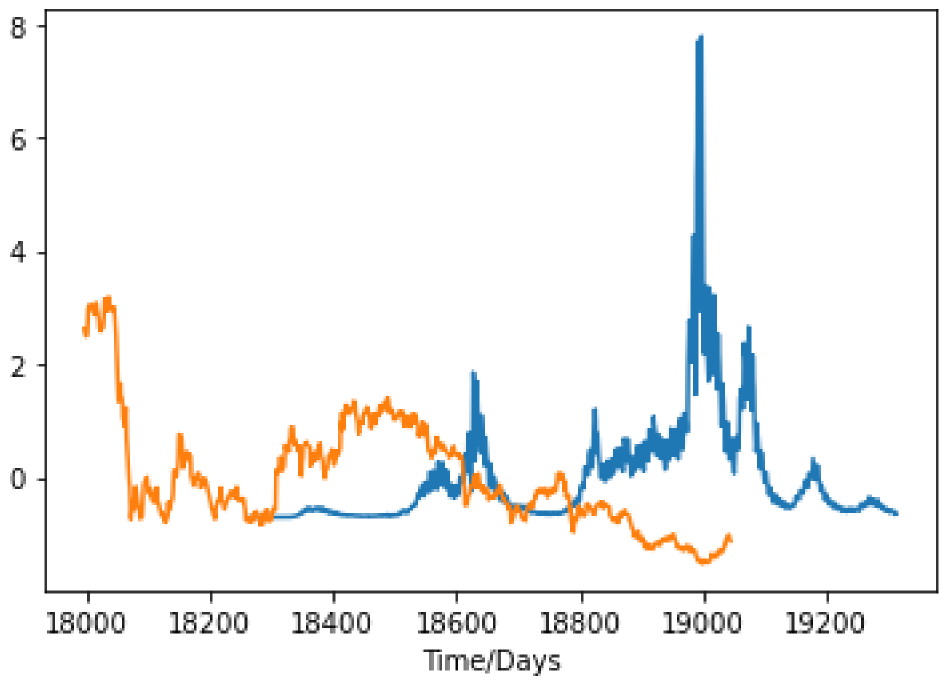

Figure 12: UK cases (blue) with Easyjet share price (orange) (with mean taken away and divided by standard deviation for normalisation) plotted with the 269.7 day offset suggested by ICCF

3.4.2 Tesco

The UK retail sector was also heavily impacted by the Covid outbreak leading to a fall of 26 percent in turnover in the first half of 2020 [7]. One would therefore expect a strong negative correlation between Covid cases and Tesco's share price.

Figure 13 (3.4.2) exhibits the best correlation occurs when the offset is negative. This offset means the Tesco's share price leads the Covid cases, meaning it's not possible to use the number of Covid cases to predict the movement. This correlation is dominated by the significant drop in Tesco's share price seen in figure 14 (3.4.2) which is most likely due to the announcement of the lockdowns within the UK when Covid cases started rising.

The DCCF has a very broad distribution, which can be seen below in figure 15 (3.4.2). Looking at the plot of the DCCF, two different peak offsets with roughly the same correlation coefficient can be seen, which prevents the code from calculating the centroid as it is unable to distinguish between the peaks and determine the largest peak. To address this issue, I isolated the two peaks to enable the centroid to be found.

This first peak identified by the DCCF (figure 16 (3.4.2)) is very close to the peak identified by the ICCF, which indicates that it is the best correlation as it was identified by both CCFs.
Similar to comparing the Covid cases to Easyjet, the other peak identified by the DCCF (figure 17 (3.4.2)) showed alignment of the offset between the tallest point and lowest point. A plot of this can be seen in figure 26 (A).

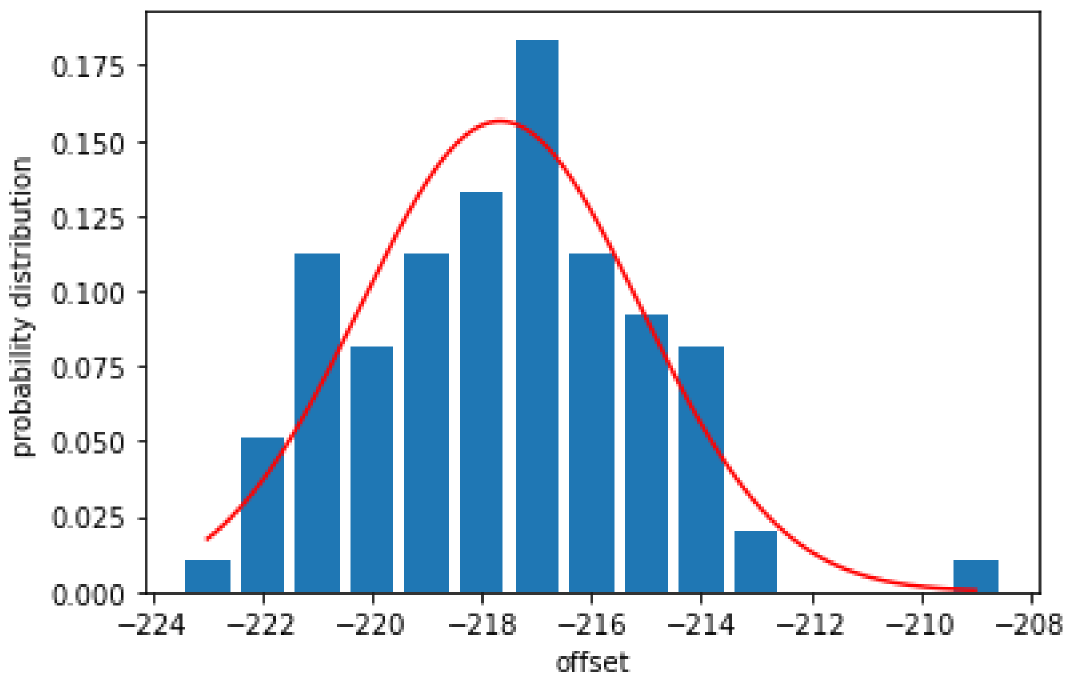

Figure 13: Probability distribution of UK cases against Tesco's share price using ICCF, peak lag of Tesco$= -217.7 \pm 0.3$ and peak correlation coefficient $= -0.59 \pm 0.001$

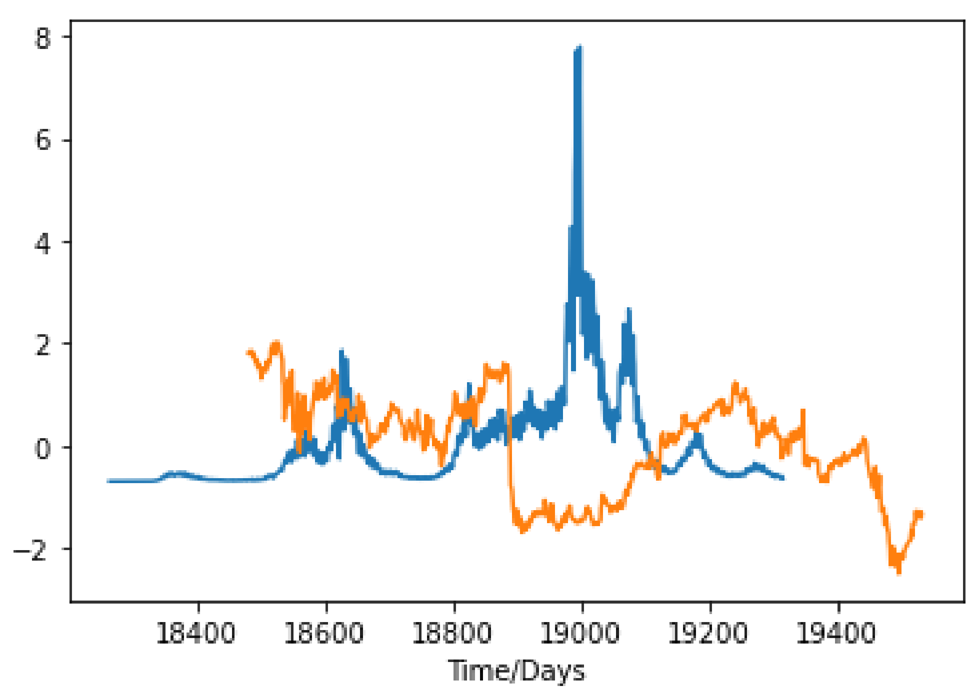

Figure 14: UK cases(blue) with Tesco's share price (orange) (with mean taken away and divided by standard deviation for normalisation) plotted with the -217.7 day offset suggested by ICCF

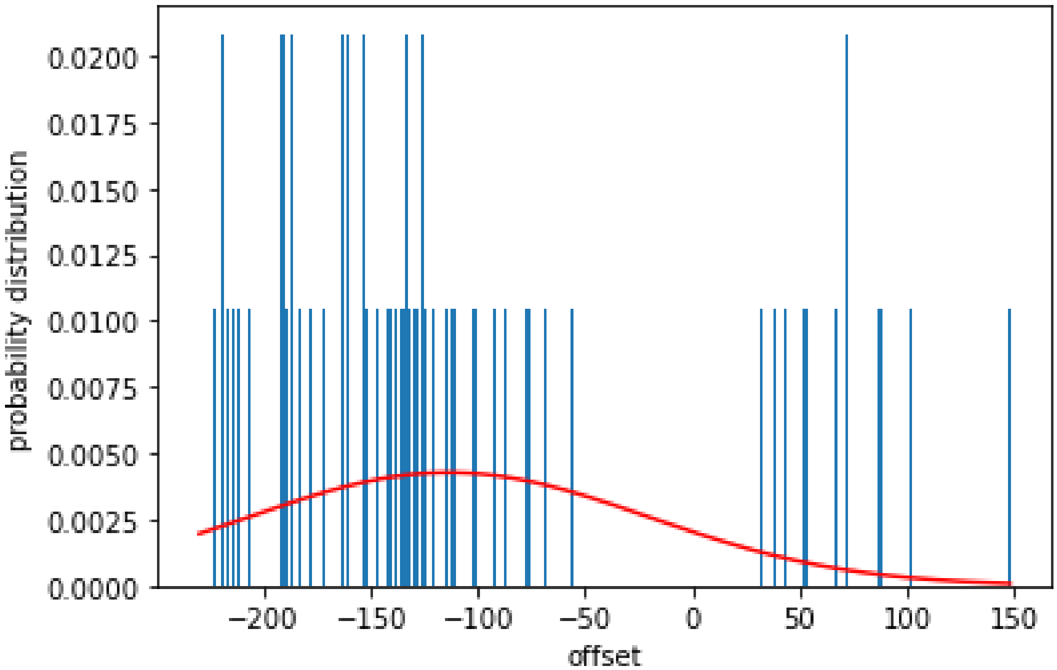

Figure 15: Probability distribution of UK cases against Tesco's share price using DCCF

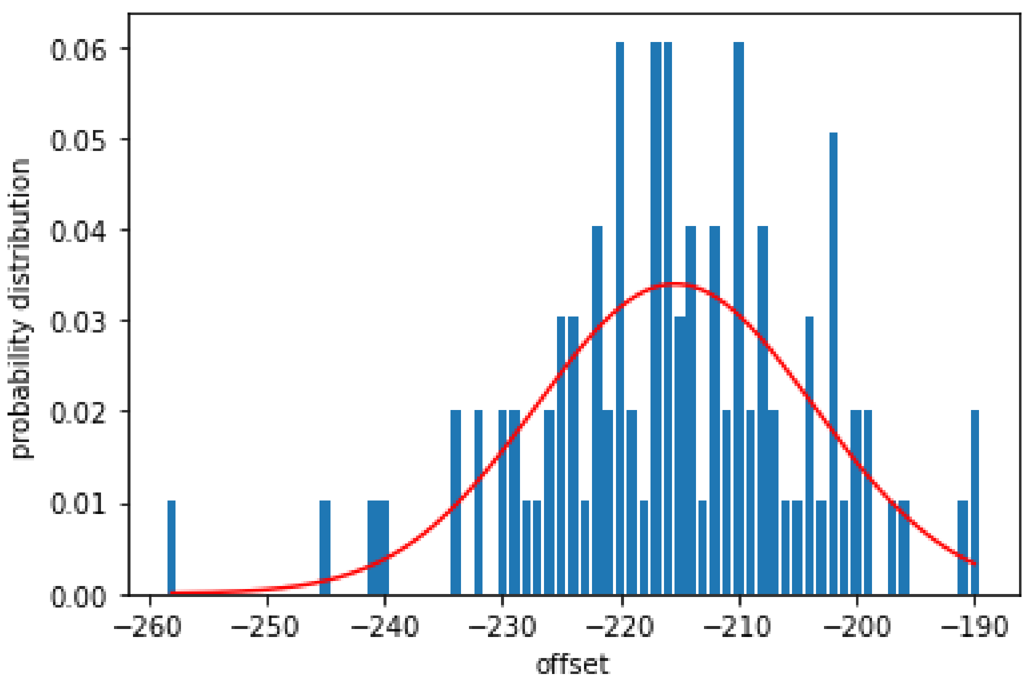

Figure 16: Probability distribution of UK cases against Tesco's share price using DCCF, peak lag of Tesco= -215.4 ± 1.2 and peak correlation coefficient = -0.547 ± 0.005

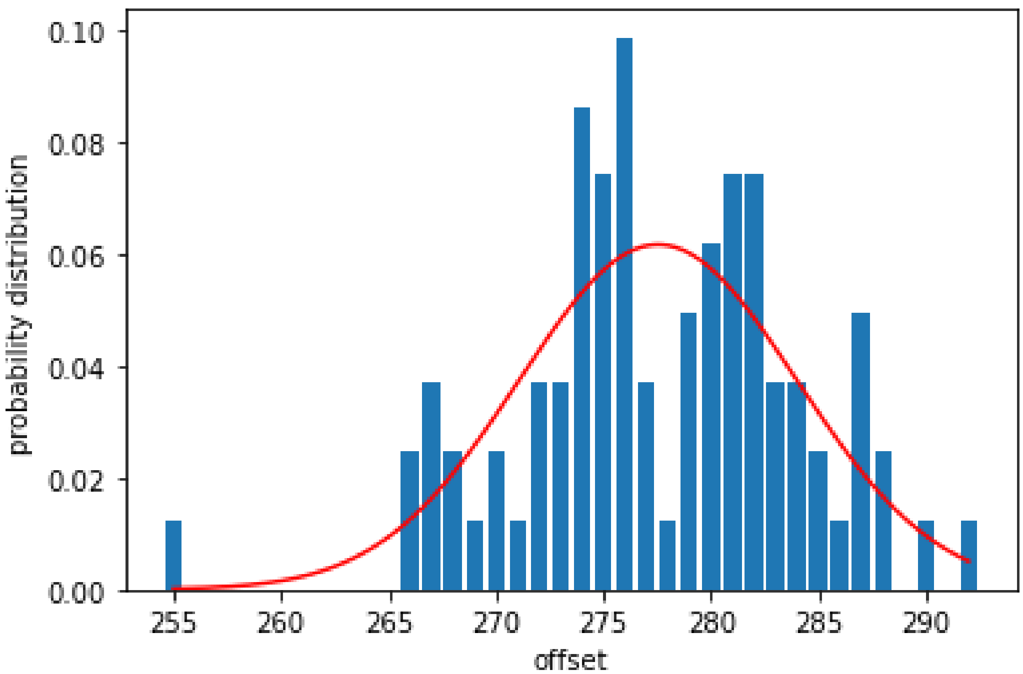

Figure 17: Probability distribution of UK cases against Tesco's share price using DCCF, peak lag of Tesco= 277.5 ± 0.7 and peak correlation coefficient = -0.455 ± 0.009

3.4.3 Whitbread

Accommodation and Food was the most affected sector by the Covid outbreak [7], so a strong negative correlation between the number of UK cases and Whitbread's share price would be expected.

However, the the ICCF and DCCF both only found a positive correlation coefficient, with a peak at -303.8, which would indicate that Whitbread's share price would rise inline with an increase in the number of Covid cases at an offset of -303.8 despite the negative impact of Covid on revenue across the overall sector.

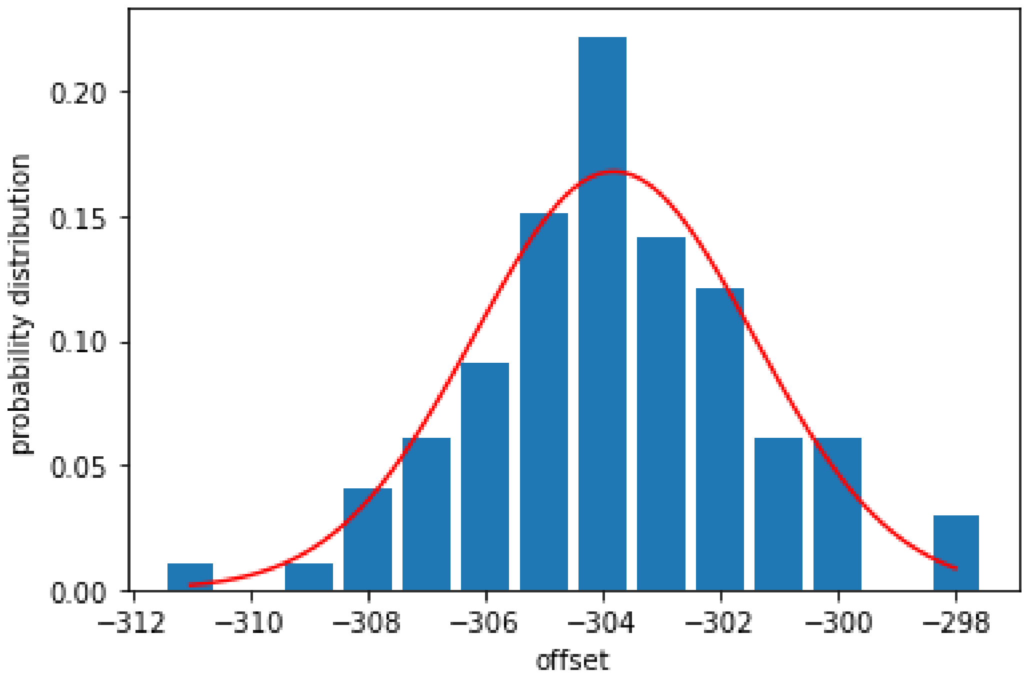

Figure 18: Probability distribution of UK cases against Whitbread's share price using ICCF, peak lag of Whitbread$= -303.8 \pm 0.2$ and peak correlation coefficient $= 0.46 \pm 0.002$

Reviewing the shifted data, it can be seen that the two lowest points of the data have been matched with each other but there's little other evidence of a correlation.

Applying a zero offset in figure 19 (3.4.3), no significant features of Whitbread's share price can be observed after the initial drop and subsequent rise in price, which suggests that it wasn't particularly affected by the increased number of Covid cases but rather the consequent lockdowns imposed by the UK government leading to one sharp drop with no further correlation.

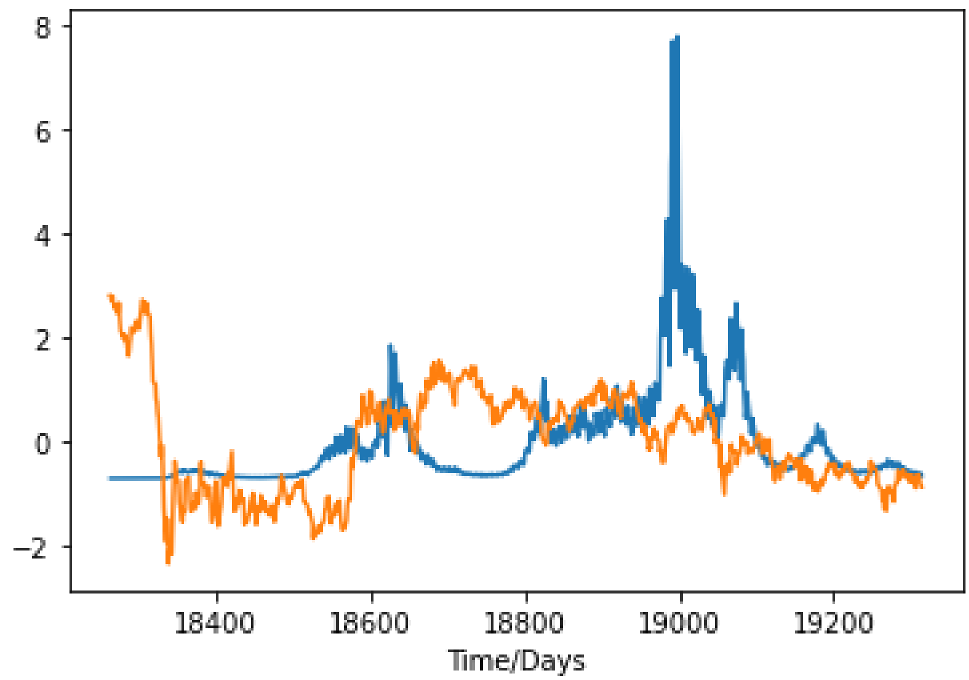

Figure 19: UK cases (blue) with Whitbread's share price (orange) (with mean taken away and divided by standard deviation for normalisation) plotted with 0 day offset

3.4.4 Cineworld

Cineworld is a large UK cinema chain that was heavily affected by Covid in the way as many leisure activities. The plot of the ICCF shown in figure 20 (3.4.4) suggests that the split between −5 and 17 was caused by a very broad and quite shallow trough indicated by the correlation coefficient of −0.361. Consequently, the minimum correlation coefficient would have varied significantly when subject to different simulations.

Looking at figure 27 (A), the correlation coefficient for the DCCF is even lower and the peak is even broader, although this will also be due to the fact that the DCCF generally has a broader probability distribution. Like before the DCCF found the best correlation when the lowest point and the highest point were aligned, I ignored it as it doesn't represent the whole data set.

The ICCF offset value appears to be the more reliable method from the previous tests as it has the higher correlation coefficient. This initial shift indicates that Covid data does lead the share price and it would be possible to predict the stock from the daily cases as the share price increases after the first peak. However, the share price starts to trend downwards after the initial increase and continues in spite of the subsequent increase in the number of Covid cases. Consequently, it would be difficult to use Covid data to reliably predict the movement of the share price despite the initial correlation.

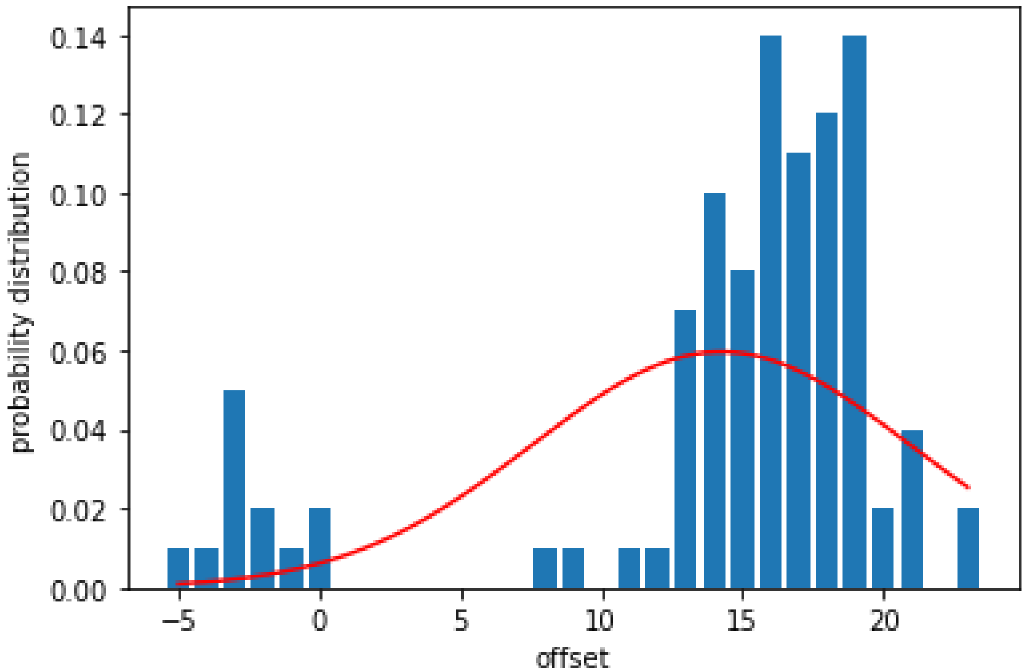

Figure 20: Probability distribution of UK cases against Cineworld share price using ICCF, peak lag of Cineworld= 14.3 ± 0.7 and peak correlation coefficient = -0.361 ± 0.002

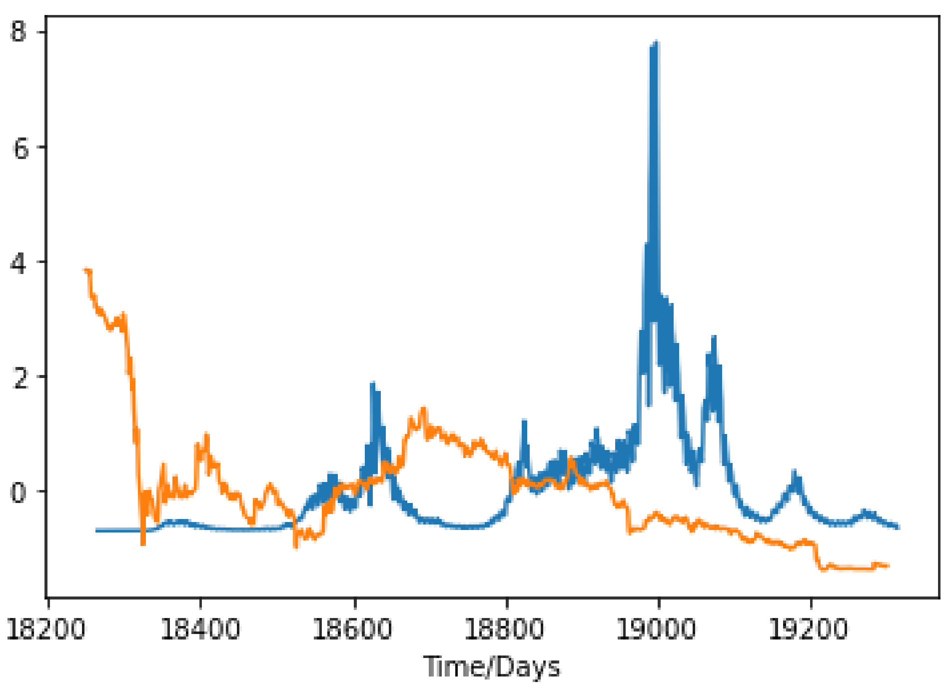

Figure 21: UK cases (blue) with Cineworld's share price (orange) (with mean taken away and divided by standard deviation for normalisation) plotted with 14 day offset suggested by ICCF

3.4.5 Comparing European shares

After confirming that it would be possible to find correlations that would enable investors to use Covid data to predict the future share price of some companies, I wondered if the logic could be applied to other regions. Given that Easyjet was the only company I tested whose share price appeared to demonstrate a strong enough correlation to support a reliable prediction, I selected Lufthansa for further comparison as it's the largest airline group in Europe.

As seen before, the DCCF only matched the the peak with the lowest point. However, the ICCF appeared to find a reasonably strong correlation. With a correlation coefficient of $= -0.537 \pm 0.001$ and

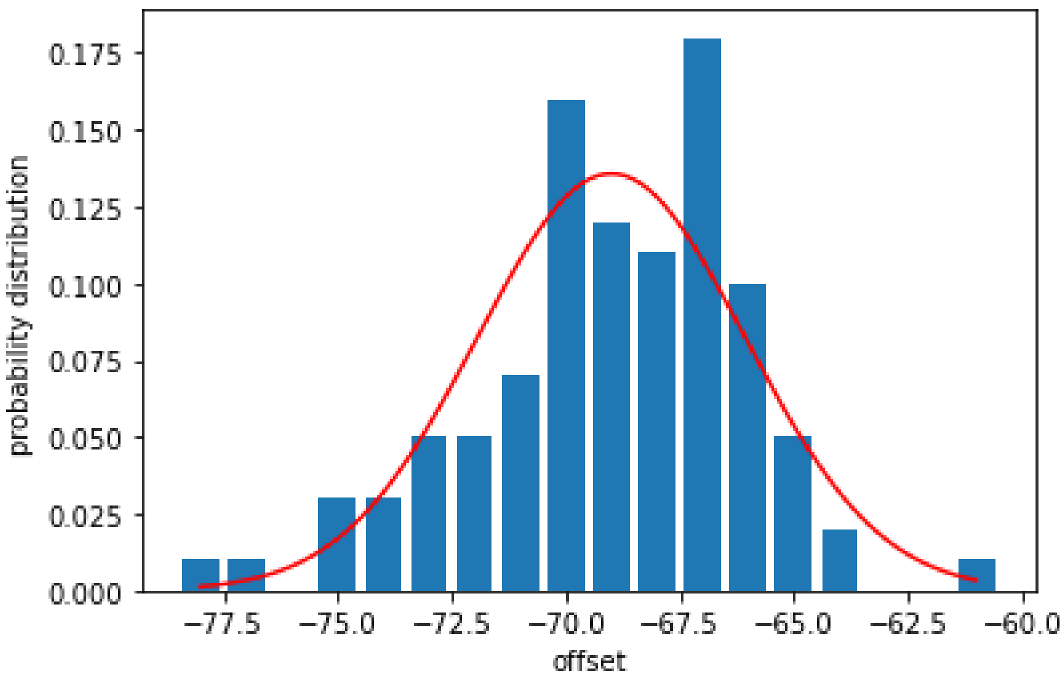

Figure 22: Probability distribution of European cases against Lufthansa's share price using ICCF, peak lag of Lufthansa$= -69.0 \pm 0.3$ and peak correlation coefficient $= -0.537 \pm 0.001$

a fairly narrow distribution, as shown in figure 22 (3.4.5). This indicates the possibility of European Covid cases having an effect on Lufthansa's share price.

However, the offset data in figure 23 (3.4.5) indicates that most of the contribution towards the high value of the correlation coefficient is generated from the large peak and the prolonged dip in the share price. Before the peak in the Covid cases, the data doesn't seem especially correlated. This supports the conclusion of the paper "Impact of COVID-19 on stock market efficiency: Evidence from developed countries" [5], which found that regions were affected by Covid in different ways and European shares were less affected by Covid cases than those in the UK. Additionally, it wouldn't be possible to use Covid data to reliably predict movements in Lufthansa's share price as it actually leads the Covid cases.

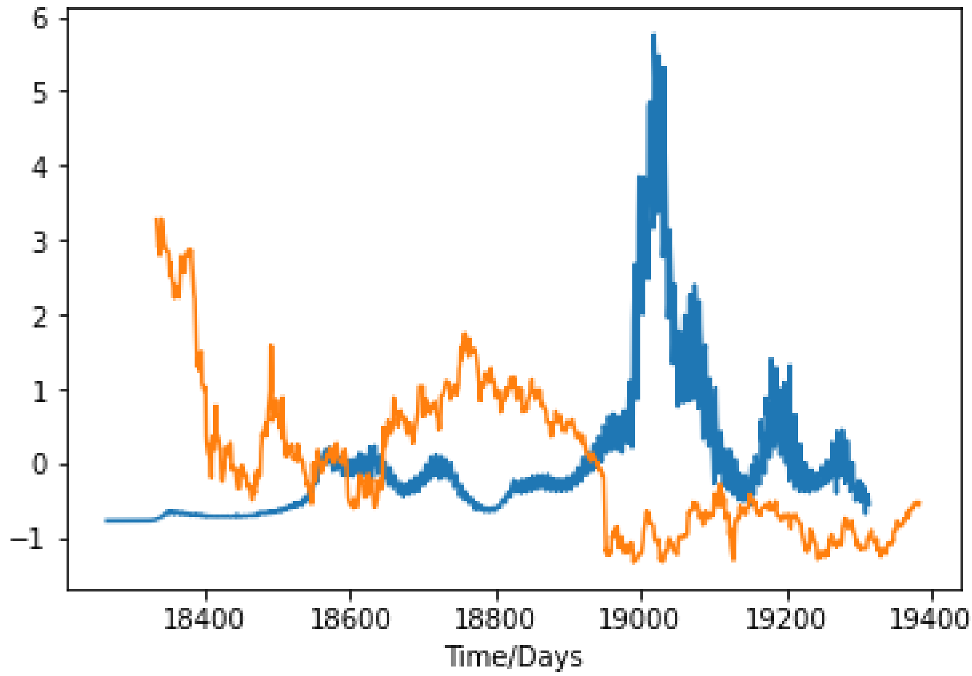

Figure 23: Europe cases (blue) with Lufthansa's share price (orange) (with mean taken away and divided by standard deviation for normalisation) plotted with −69 day offset suggested by ICCF

3.5 Trying to fix finding correlation with the DCCF

After observing that the DCCF struggled to find a correlation that wasn't dictated by the highest and lowest points of each data set for almost all of the stocks I tested, I attempted to determine a reason why.

Firstly, I tried reducing the range of offsets tested by the DCCF so that it excluded offsets where the highest and lowest points were aligned. This in theory would demonstrate whether the DCCF was finding a correlation that matched the ICCF but wasn't being found because the correlation coefficient of the alignment of the highest and lowest points was higher. However, when attempted the results didn't vary from the ones above.

A possible explanation could be the fact that 30 percent of the data points were deleted during the simulation, leading to fewer points and even more weighting being put on the larger points, causing them to dominate over others. However, I saw no change in the results after I adjusted the amount of data being deleted when doing the simulations.

I was unable to delete the spike in the Covid data as it occurs in the middle of the data set. It would also involve removing a significant proportion of the data points, which doesn't seem to be a suitable option. Another possible solution would be to use the local mean when calculating individual DCCF values for each point and reduce the distance from the mean (reducing the peaks weighting). However, the spike in Covid cases is so sharp that this is unlikely to have much of an effect because the points would still be much further away from the local mean than any of the other points due to the gradient of the spike. Similarly, detrending could be used to remove any underlying trends in the share price data which, like the local mean, would move the points closer to the mean. Nonetheless, it may be worthy of further investigation, with additional time.

4 Conclusion

This project aimed to find if there was a correlation present between the daily reported Covid cases and the performance of share prices in different industry sectors using the Discrete Cross-Correlation Function and the Interpolated Cross-Correlation Function to determine if they could be used to generate profit during the Covid pandemic.

Before looking at the specific values of offsets calculated, it is important to discuss the different results between the DCCF and the ICCF despite both methods appearing to do the same thing. Although the auto-correlation verified that the code was working correctly at the start of the investigation, the DCCF and ICCF haven't been producing the expected same results. Only when changing parameters did it find a correlation that was within the error bars of the ICCF. It appears that the DCCF is more sensitive to large spikes in the data compared to the ICCF which seems to find correlations in general trends, a likely consequence of the interpolation of the data smoothing out any spikes. While this does not mean the DCCF was incorrect, the correlations weren't useful for finding correlations which take into account the majority of the data, which is necessary to generate reliable predictions.

Once the ICCF had shown to give more suitable offsets, there were only two companies whose share prices trailed the Covid cases: Easyjet with an offset of 20 days and Cineworld with and offset of 14 days. Tesco's shares appeared to move before the Covid cases with an offset of -218 days.

Although both Easyjet and Cineworld's shares started a downward trend as the second spike of cases start to form, the trend remained almost unchanged until the end of the 18^{th} November 2022, long after press coverage and Covid cases had reduced, but it is not possible to determine when Covid's influence is replaced by alternative dynamics.

Further, the comparison between Easyjet and Lufthansa to explore possible regional variations showed that not only did Lufthansa not produce the same offset with European cases as Easyjet did UK cases but Lufthansa lead the European cases by -69 days. This suggests each shares in each region are affected by Covid cases by different magnitudes.

In conclusion, it may have been possible to use Covid data to vaguely predict the movement certain shares, any predictions are unlikely to carry across to other companies as each responded differently and therefore need to be tested individually. This would have been a time consuming exercise that may not have been possible to complete in time before the information had been factored into the share price. Likewise, there is no definitive way of knowing whether a correlation could be specifically attributed to the impact of Covid or caused by an unrelated influence.

Bibliography

[1] yahoo finance. In: (). URL: https://uk.finance.yahoo.com/.

[2] Laurence Fletcher. "Top hedge funds reap biggest gains in a decade during pandemic". In: *The Finacial Times* (). URL: https://www.ft.com/content/2d106227-d6e7-4ea2-9772-abb0da694e70. (accessed 27/12/2022).

[3] By Steven L. Jones and Jeffry M. Netter. "Efficient Capital Markets". In: (). URL: https://www.econlib.org/library/Enc/EfficientCapitalMarkets.html. (accessed 2/1/2023).

[4] World Health Organisation. "Daily cases and deaths by date reported to WHO". In: (). URL: https://covid19.who.int/data. (accessed 25/11/2022).

[5] Oktay Ozkan. "Impact of COVID-19 on stock market efficiency: Evidence from developed countries". In: (2021). URL: https://www.sciencedirect.com/science/article/pii/S0275531921000660. (accessed 24/12/2022).

[6] Bradley M. Peterson et al. "On Uncertainties in Cross-Correlation Lags and the Reality of Wavelength-dependent Continuum Lags in Active Galactic Nuclei". In: *Publications of the Astronomical Society of the Pacific* (). URL: http://www.jstor.org/stable/10.1086/316177. (accessed 26/12/2022).

[7] Ginevra Okoye Sumit Dey-Chowdhury Muhammed Khaliq and Libby Abrahams. "Effects of the coronavirus (COVID-19) pandemic on 'high-contact' industries". In: (2022). URL: https://www.ons.gov.uk/economy/grossvalueaddedgva/articles/%20effectsofth%20ecoronaviruscovid19pandemiconhigh 2022-05-06#toc. (accessed 24/12/2022).

[8] Russel J. White and Bradley M. Peterson. "Comments on Cross-Correlation Methodology in Variability Studies of Active Galactic Nuclei". In: *Publications of the Astronomical Society of the Pacific* (). URL: http://www.jstor.org/stable/40680443. (accessed 26/12/2022).

A Appendix

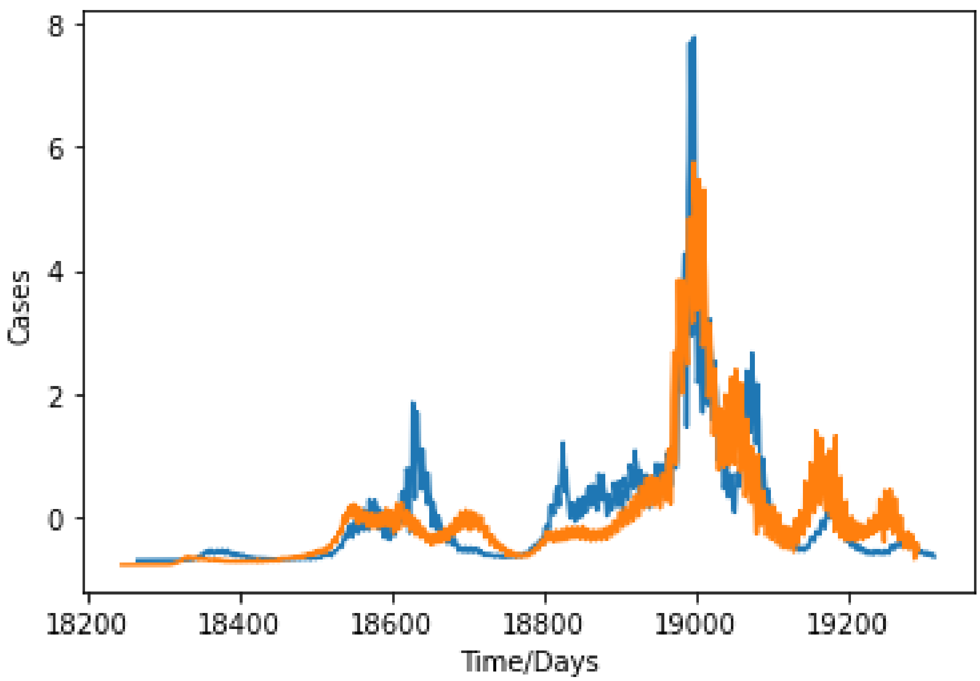

Figure 24: Euro (orange) and UK (blue) Covid cases (with mean taken away and divided by standard deviation for normalisation) plotted with the 21 day offset suggested by ICCF and DCCF

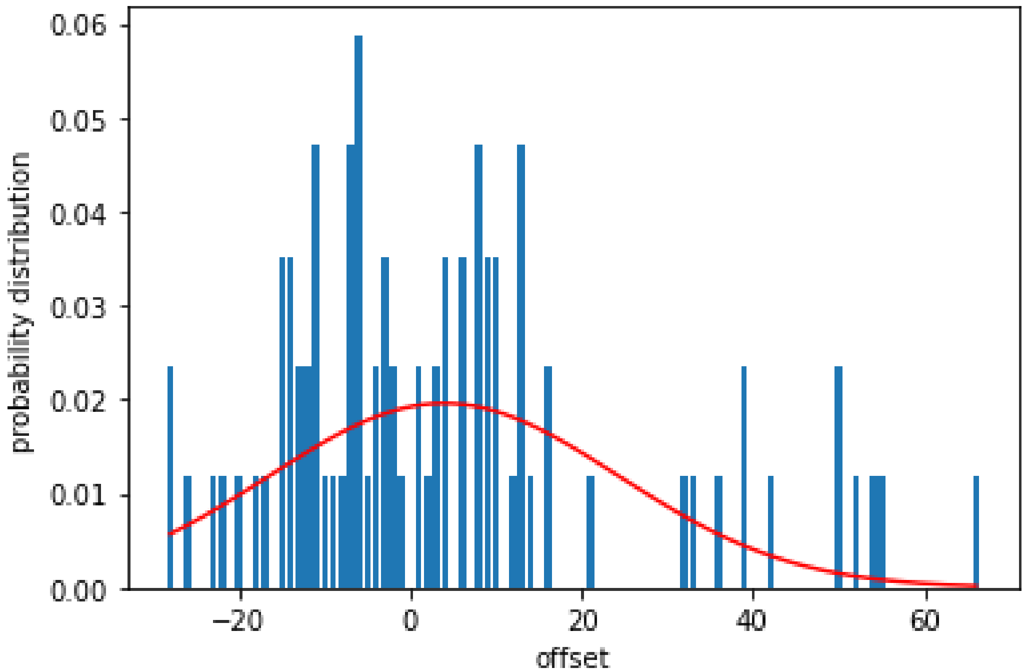

Figure 25: Probability distribution of UK cases against EasyJet's share price using the DCCF, peak lag of Easyjet= 3.9 ± 2.2 and peak correlation coefficient $= -0.145 \pm 0.005$

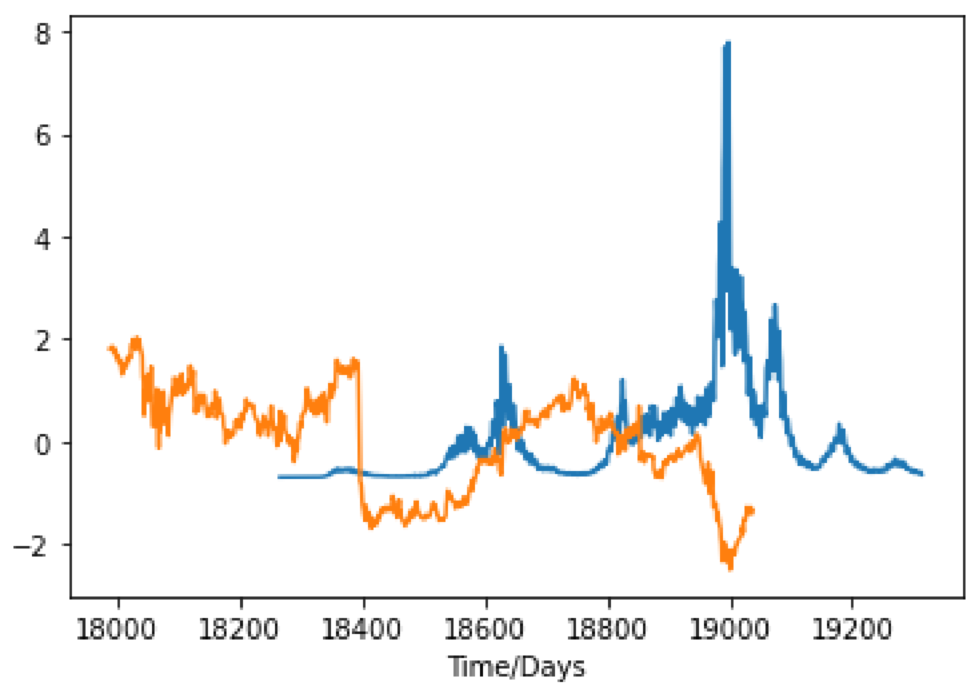

Figure 26: UK (blue) cases with Tesco's share price (orange) (with mean taken away and divided by standard deviation for normalisation) plotted with the 277.5 day offset suggested by DCCF

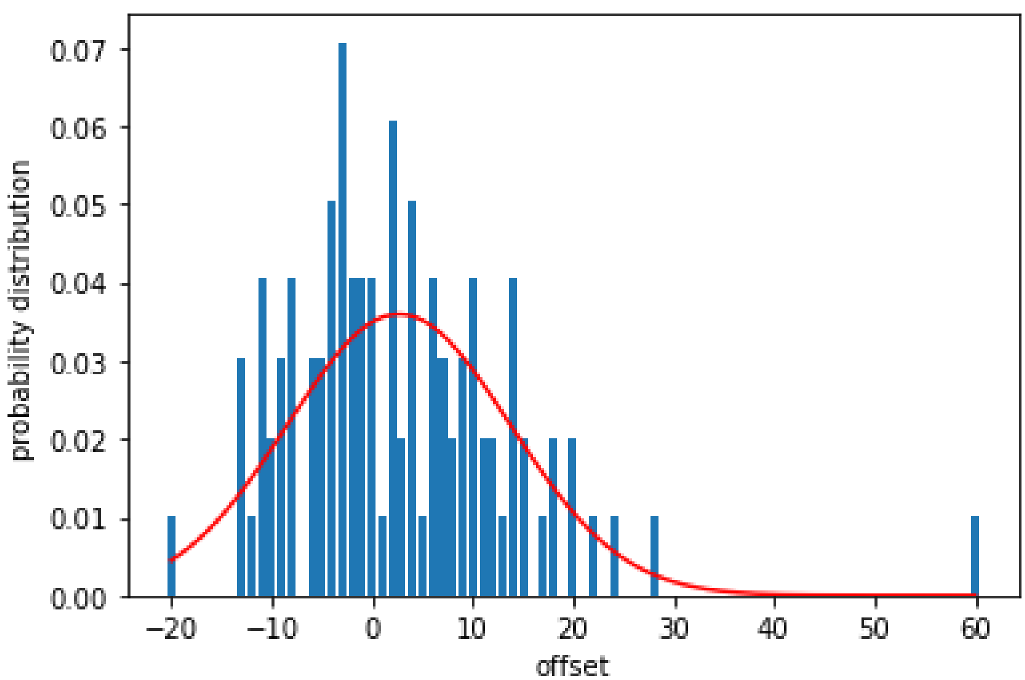

Figure 27: Probability distribution of UK cases against Cineworld's share price using DCCF, peak lag of Cineworld= 2.7 ± 1.1 and peak correlation coefficient $= -0.196 \pm 0.004$