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

James E. McDonald

January 8, 2023

1 Introduction to Cross-Correlation Functions

This project set out to determine whether it would have been possible to use two types of Cross Correlation Function (CCF) - Interpolated CCF (ICCF) and Discrete CCF (DCCF) - to reliably predict the movement of share prices [1] from the daily Covid-19 (Covid) cases [2]. The CCFs measure how well correlated two sets of data are by producing a number, called the Correlation Coefficient, between 1 and -1, where a coefficient of 1 indicates the two sets of data are exactly the same (correlated), -1 indicates the two sets of data are exactly opposite (anti-correlated) and 0 indicates there's no resemblance between the two (no correlation).

The CCFs generate a correlation coefficient for different offsets of the data, meaning one set of data is shifted by τ days. The offset with the highest or lowest correlation coefficient reveals the point where the two data sets are most correlated or anti-correlated, which can then be interpreted as how much one data set lags behind the other. Consequently, if a share price has a correlation coefficient that is high or low enough to indicate a sufficient dependence on the Covid cases and its offset implies it lags behind the Covid cases, one could reasonably predict whether the share price will move up or down based on the change in the number of Covid cases.

The stock market is closed on the weekends and bank holidays leading to missing data points on these days, the ICCF interpolates over the missing data points by adding points to make a straight line and then adding noise to imitate the noise in the share price. The DCCF however doesn't interpolate over the data leading to it having less data points to calculate the Correlation values.

To filter out noise and find errors a method of simulation is used [3] to delete 30 percent of the data randomly and calculating the ICCF and DCCF in the hope of deleting potential out lying points. This is done 100 times, creating a probability distribution of a peak offset being calculated, from which the offset with the highest probability of being calculated is taken as the best offset and the width of the distribution indicates the error.

2 Findings

2.1 ICCF compared to DCCF

Unexpectedly, the ICCF and DCCF did not agree on the best offset for any of the share prices tested against the number of Covid cases. When the share prices were plotted against the Covid cases with the offset generated by the DCCF, it was obvious it had found the best correlation where the highest point of the Covid cases and the lowest point of the share price were aligned.

This is most likely due to the DCCF having fewer data points than the ICCF because it doesn't interpolate, leading to the correlation coefficients for each data point having a higher weighting on the average for the correlation coefficient for each offset. This causes the correlation coefficient to be much higher when the highest and lowest points are aligned because the spike in the Covid cases is so

significant.

Although the two methods don't agree, it doesn't mean one method is correct and the other incorrect. The DCCF is much better at correlating extremes with each other, while the ICCF generates an offset where the majority of the data points are correlated. For this project the ICCF is a more suitable method for predicting the movements of shares due to the fact that an offset that considers the majority of points is needed.

2.2 Easyjet

Easyjet had the best anti-correlation of all of the companies I tested against the UK Covid cases, implying that the travel industry was most affected by the number of Covid cases.

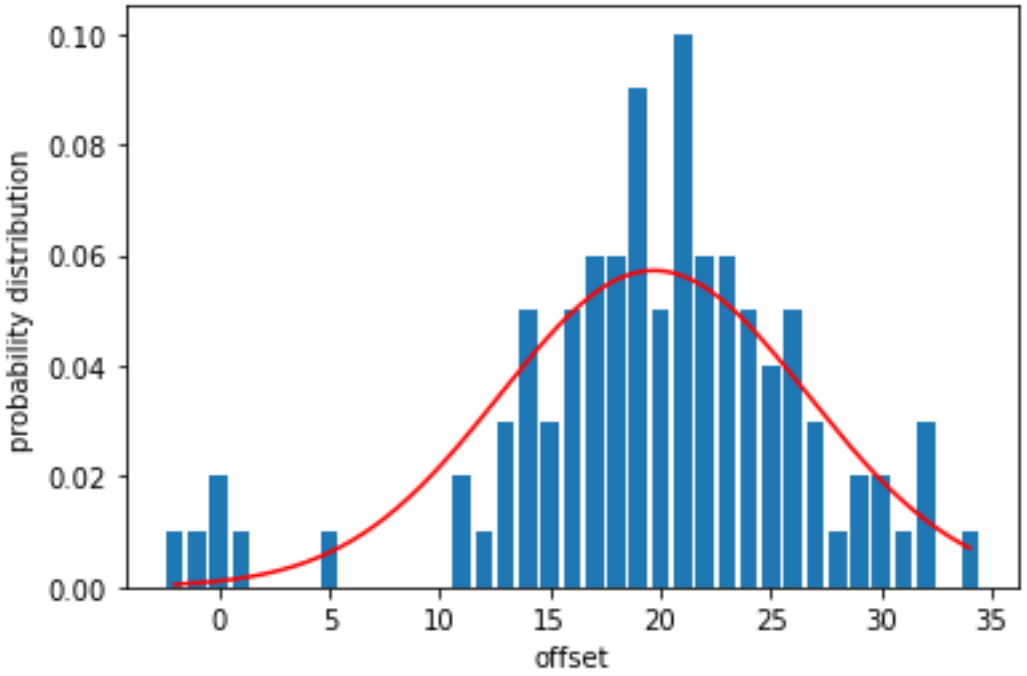

Figure 1: Probability distribution of UK cases against Easyjet's share rpice using ICCF, peak lag of Easyjet $= 19.7 \pm 0.7$ and peak correlation coefficient $= -0.339 \pm 0.002$

Figure 1 (2.2) shows that, although the correlation coefficient isn't very large, the probability distribution is well-formed and quite narrow implying that little noise contributed to this correlation. Plotting the offset share price against the Covid data in figure 2 (2.2) to check the correlation appears to be reasonable indicates that the share price fairly reliably moves in the opposite direction to the Covid cases.

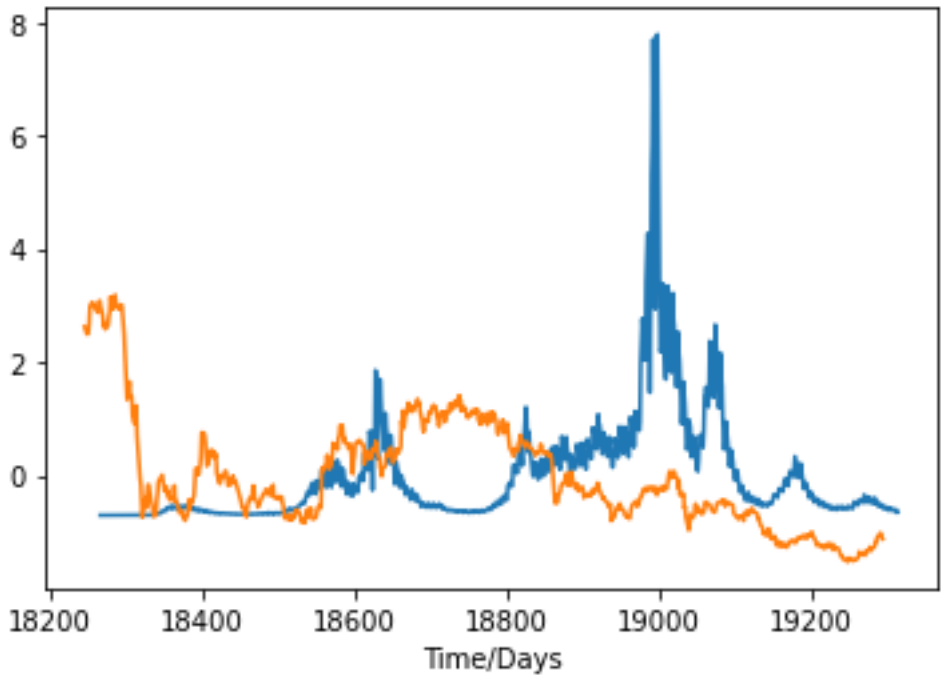

Figure 2: UK cases with Easyjet's share price (with mean taken away and divided by standard deviation for normalisation) plotted with the 20 day offset suggested by ICCF

2.3 Lufthansa

The project also included a test of Lufthansa to determine whether the offset produced by Easyjet and UK Covid cases was transferable to Lufthansa and European Covid cases.
The probability distribution produced in Figure 3 2.3 indicates Covid cases trails Lufthansa's share price by 69.0 ± 0.3, which illustrates that not only is the degree of shift different between the two airlines but Lufthansa's share price leads Covid cases while Easyjet's share price lags behind. This indicates the two companies' share prices didn't react in the same way to Covid cases despite being in the same industry.

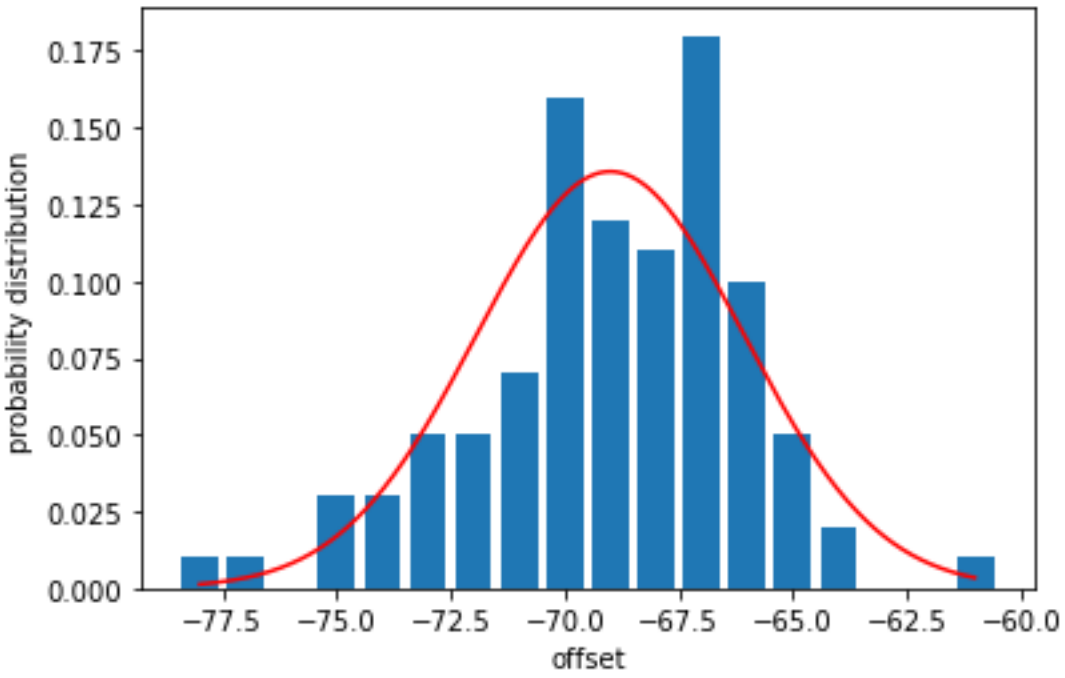

Figure 3: Probability distribution of European cases against Lufthansa's share price using ICCF, peak lag of Lufthansa $= -69.0 \pm 0.3$ and peak correlation coefficient $= -0.537 \pm 0.001$

2.4 Other findings

From the companies I tested that didn't produce a significant correlation, a recurring feature was a large initial drop at the beginning on the plots of the time series, suggesting that the share prices reacted to the initial lockdowns imposed by the UK governments but weren't necessarily affected by the daily cases.

Also, looking at the Easyjet share price plotted against the shifted Covid cases in Figure 2, it's apparent that the influence of Covid cases on the Easyjet share price diminishes over time as Easyjet's share price fails to recover as the number of Covid cases falls after the second spike. Finding the point at which this influence stops, and therefore the point at which one could no longer predict the movement of the share price from Covid cases, would be difficult. Consequently, predicting the share price would be unreliable.

3 Conclusion

This project found that Easyjet's share price lagged behind the daily UK Covid cases by 20 days, meaning it would have theoretically been possible to predict the movements of Easyjet's share price from Covid cases. However, the reliability of doing this would require further investigation to build confidence as the dependence of Easyjet's share price on Covid cases decreases without warning. As well as this, the offset produced for a company isn't transferable to another as demonstrated by Lufthansa and Easyjet meaning each company would need to be tested independently.

Bibliography

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

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

[3] 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).