Lisbon School of Economics & Management
Universidade de Lisboa

MASTER OF SCIENCE IN FINANCE

MASTER´S FINAL WORK

DISSERTATION

COVID-19 AND STOCK MARKET VOLATILITY: A CLUSTERING APPROACH FOR S&P 500 INDUSTRY INDICES

FRANCISCO LÚCIO

DECEMBER-2022

MASTER OF SCIENCE IN
FINANCE

MASTER´S FINAL WORK
DISSERTATION

COVID-19 AND STOCK MARKET VOLATILITY: A CLUSTERING APPROACH FOR S&P 500 INDUSTRY INDICES

FRANCISCO LÚCIO

SUPERVISION:

JORGE CAIADO

DECEMBER-2022

GLOSSARY

AR – Autoregressive model

ARCH – Autoregressive conditional heteroskedastic model

CAPM – Capital Asset Pricing Model

EPU – Global Economic Policy Uncertainty Index

EUA – Estados Unidos da América

GARCH – Generalized Autoregressive Conditional Heteroskedastic model

GICS – Global Industry Classification Standard

JEL – Journal of Economic Literature

MAPE – Mean Absolute Percentage Error

MSCI – Morgan Stanley Capital International

PCA – Principal Component Analysis

S&P – Standard & Poor's

TGARCH – Threshold Generalized Autoregressive Conditional Heteroskedastic Model

U.S. – United States

VIX – Chicago Board Options Exchange Volatility Index

ABSTRACT, KEYWORDS AND JEL CODES

The COVID-19 pandemic was the infectious disease outbreak that has had the strongest impact on the U.S. stock market. In this dissertation, we study how this impact affected some of the conditional volatilities of S&P 500 industries, using a new model feature-based clustering method on a fitted threshold generalised autoregressive conditional heteroscedasticity (TGARCH) specification. Rather than using the estimated model parameters to compute a distance matrix for the stock indices, which cannot capture all the information about the dependence of the time-varying variance, we suggest using a distance based on the autocorrelations of the estimated conditional volatilities. Both hierarchical (complete linkage) and non-hierarchical (k-means) unsupervised machine learning algorithms are used to assign the set of industries into clusters. The results show a clear change in the composition of each cluster between the period before the first U.S. COVID-19 case and the period during the pandemic, leading to the conclusion that the similarities or distances between industries underwent a significant change, with the industries most affected by the pandemic being Hotels, Consumer Durable & Apparel, Automobile, and Airlines. It was also made an analysis regarding the forecast accuracy of simple asymmetric GARCH models applied to S&P 500 industries and use the model forecast errors for different horizons to calculate a distance matrix for the stock indices. A hierarchical clustering algorithm is used to assign the set of industries into clusters. We found homogeneous clusters of industries in terms of the impact of COVID-19 on US stock market volatility. The industries most affected by the pandemic and with less accurate stock market prediction (Hotels, Resorts & Cruise Lines, Airline, Apparel, Accessories & Luxury Goods, and Automobile) are separated in Euclidean distance from those industries that were less impacted by COVID-19 and which had more accurate forecasting (Pharmaceuticals, Internet & Direct Marketing Retail, Data Processing & Outsourcing Services, and Movies & Entertainment).

JEL: C32, C38, G11, G170.

Keywords: Autocorrelation; Cluster analysis; COVID-19; Threshold GARCH model; Unsupervised machine learning; S&P 500; Volatility; Forecast accuracy

RESUMO, PALAVRAS-CHAVE E CÓDIGOS JEL

A pandemia do COVID-19 foi o surto com maior impacto de sempre no mercado de ações dos EUA. Nesta dissertação, estudamos como este impacto afetou algumas das volatilidades condicionadas das indústrias do S&P 500, usando métodos de clustering aplicados numa especificação de um modelo derivado (*threshold)* do generalizado autorregressivo condicional e heterocedástico (TGARCH). OS parâmetros do modelo estimado não capturam toda a informação acerca da dependência da variância não constante ao longo do tempo pelo que usamos distância baseada nas autocorrelações das volatilidades condicionais, para calcular a matriz de distância para os índices de ações. Ambos os algorítmos hierárquicos (*complete linkage)* e não-hierárquicos (k-means), técnicas de *unsupervised machine learning,* são usados para agrupar conjuntos de indústrias em *clusters*. Os resultados mostram uma clara mudança na composição de cada cluster entre o período anterior ao primeiro caso de COVID-19 nos EUA e o período durante a pandemia. As semelhanças ou distâncias entre as indústrias sofreram uma mudança significativa, sendo as indústrias mais afetadas pela pandemia, os Hotéis, Bens de Consumo Duráveis e Vestuário, Automóveis e Companhias Aéreas. Foi feita uma análise quanto à precisão da previsão de modelos GARCH assimétricos simples aplicados a indústrias do S&P 500. Esses erros de previsão calculados em diferentes intervalos de tempo, durante a pandemia, foram usados para calcular a matriz de distâncias entre os indices de mercado. Um algoritmo de *hierarchical clustering* foi usado para agrupar um grupo de indústrias em clusters. Obtivemos clusters homogéneos em termos do impacto do COVID-19 na volatilidade do mercado de ações nos EUA. As indústrias mais afectadas pela pandemia, nas quais os modelos preditivos se mostraram menos precisos em termos preditivos (*Hotels, Resorts & Cruise Lines, Airline, Apparel, Acessories & Luxury Goods, and Automobile*) estão separadas por uma distância Euclideana das indústrias que foram menos impactadas pelo COVID-19 e as previsões mais precisas (Pharmaceuticals, Internet, Data Processing, e Movies & Entertainment).

JEL: C32, C38, G11, G170.

Palavras-chave: Autocorrelação; Cluster analysis; COVID-19; Threshold GARCH model; Unsupervised machine learning; S&P 500; Volatilidade; Precisão de previsão

ACKNOWLEDGEMENTS

This dissertation was written during the most defiant, but ultimately happiest period of my life.

I wrote the most part of my thesis while I was living in Milano, studying at Bocconi University, where I attended two courses: Machine Learning and Time Series Analysis of Economic and Financial Data. I am truly passionate by machine learning and data analytics, and I am immensely grateful for this opportunity that allowed me to conciliate my master's final work with two courses that definitely contributed a lot to produce it. The fact that my dissertation was mostly produced in Italy gives it a special meaning for me and I have to thank my friends that turned this journey into an amazing life experience. Italy is beautiful, but one of the things I've learned is that paradise is not a place that we can look for, it is a feeling that sometimes urges in a moment of unmeasurable happiness, and that sticks with us forever. None of this would have possibly been the same without the support and the company of my friends that never let me feel lonely during these 5 months of my life. So, from the deep of my heart, Carolina Tapada, Giovanni Rapazzini, Jannis Stefanopulos, Luca Peselli, Matteo Dorigo, Nicolò Libertino, Simone Dall'Ara, Sofia Miranda and Sonam Berto*, grazie mille a tutti.* I also want to thank to Ana Mamede, Carolina Batista, Catarina Candeias and Patricia Lopes for all the support and for caring about me while I was far away. Now, to the people that were crucial for this adventure to happen, both my parents and my aunt, please accept my deepest gratitude.

The presentation and the process of reviewing was done while living in Dublin. For that reason, the most sincere thanks to all the friends I have met here and to everyone at Deloitte, specially in my team, who was always so supportive and comprehensive with me during this process. Really pleased.

While I was choosing the topic, I wanted to assure I had the right amount of Time Series knowledge to base my dissertation in this subject. My friend José Neves, student of Actuarial Science and my colleague during our bachelor's degree in Mathematics Applied to Economics and Management, sent me a lot of his Time Series notes and programming codes, for me to be better prepared for this challenge, which I much appreciated.

Finally, the biggest and most heartfelt thanks to my supervisor Jorge Caiado. At the beginning, as a big enthusiast by the movie industry, I had a foggy idea about studying the impact of the COVID-19 pandemic in the Movie & Entertainment industry stock market. My supervisor was the sunrise I needed to structure all those hazy ideas into something solid and innovative. He tirelessly helped me with scientific writing, new suggestions, programming code and also with time management. No words are enough to thank him, and without his help I would have struggled a lot more along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

In the 9th of January, the World Health Organization announced that an infectious coronavirus had struck Wuhan, in China. In a few weeks the virus started spreading all across the world and on the 21st of January, the first United States (U.S.) case was registered. Later, on the 11th of March, the World Health Organization declared COVID-19 to be a pandemic.

The uncertainty generated by the whole pandemic situation with regards to mortality, government policies and individual behavior affected the U.S. stock market returns and volatilties (Song, Yeon and Lee, 2021). The U.S. stock market crashed in March 2020, following the government's reaction to COVID-19, which was primarily related to the imposition of strict lock downs and closure of diverse businesses (Mazur et al. 2020).

The impact of the COVID-19 pandemic on the U.S. stock market has been studied in the scientific literature by using a number of different approaches. Baker et al. (2020) used newspaper cuttings and on-line articles to analyse the impact of news about the outbreaks of infectious disease on the stock market. Other papers centre their analysis on stock market returns (e.g. Cepoi, 2020; Ashraf, 2020). Cepoi (2020) used the quantile regression model to prove the existence of asymmetric dependencies related with COVID-19 information in the stock market, while Ashraf (2020) studied the impact of the number of confirmed COVID-19 cases and deaths on stock market returns, using panel data analysis. Other studies focus more on the COVID-19 spread impact in the stock market volatility (e.g. Albulescu, 2021; Chaudhary et al., 2020; Baek et al., 2020). Albulescu (2021) used a simple Ordinary Least Squares regession to measure the impact of the COVID-19 spread on the U.S. financial volatility. Chaudhary et al. (2020) used a GARCH model to study the influence of COVID-19 both on the return and volatilities in different stock markets. Baek et al. (2020) used a Markov-Switching AR model to investigate a possible relation between COVID-19 and a change in the volatility of the U.S. stock market. Some papers centre their analysis on measuring this impact by using volatility models as forecasting tools (Ertugrul et al. 2020). Ertugrul et al. (2020) used variants of the generalized ARCH models to study the COVID-19 pandemic impacts on the Turkish diesel market. He compared the forecasting accuracy of these different models to assess which one is the best to acess the disruptive effects of the COVID-19 spread on the Turkish diesel market. Other authors, such as, Wang et al. (2020) studied which predictors better forecast international stock markets during the COVID-19 spread. Wang et al. (2020) compares the accuracy of the predictors VIX and EPU indexes to investigate which one is best to forecast future volatility during the coronavirus pandemic. None of the studies previously cited used unsupervised learning algorithms for time series analysis. Indeed, the state of the art methodology (specifically the study of the relationship between COVID-19 and stock market volatility) does not comprise any clustering method that is capable to take into account all the available useful information about the stochastic dependence of the stock market volatilities.

In this study, we investigate how the COVID-19 pandemic affected the volatility of the U.S. stock market, with a focus on major S&P 500 industries, including Communication Services, Consumer Discretionary, Healthcare, Industrials, Energy, Financials, and Information Technology. We describe 11 S&P 500 industries in terms of their empirical properties and stylised facts and employ a cluster analysis approach to identify similarities and dissimiarties among industries before and during the spread of the COVID-19 pandemic. Then, we evaluate the forecast accuracy of parcimonious threshold generalized autoregressive conditonally heteroscedasticity (TGARCH) models to predict those sub-industry indices. For this purpose, we calculated the mean absolute percentage errors (MAPE) of multi-step-ahead forecasts for 23 different cutoff points during COVID-19 pandemic. Finally, we employ a model-based clustering approach to identify common dynamic features among industries during pandemic. This method uses the Euclidean distances between forecast errors of fitted TGARCH models. We believe that if the forecast errors of two or more industries have a high degree of similarity, then models generating these forecast errors should also have a high degree of similarity.

Cluster analysis of financial time series data has been studied in past research using both time and frequency domain-based methods (see Caiado and Crato, 2007, Otranto, 2008, Otranto, 2010, Caiado and Crato, 2010, Bastos and Caiado, 2014, Caiado et al., 2015, D'Urso et al., 2016, Maharaj et al., 2019, Caiado et al., 2020, and Bastos and Caiado, 2021). In one of the first relevant contributions to this research topic (Caiado and Crato, 2010), the authors introduced a metric for heteroskedastic time series based on the estimated parameters of the threshold generalised autoregressive conditional heteroscedasticity (TGARCH) model. Since the parametric features of the TGARCH representation (ARCH, GARCH and leverage terms) are not associated with all the autocorrelation lags of the time-varying volatility, we suggest using a distance based on the autocorrelations of the estimated conditional variances that can be weighted with their estimated variances and covariances.

For this purpose, we estimate the conditional variances of each one of these industries indices from a parsimonious TGARCH model and compare the time-varying volatility estimated parameters of its representation before and during the COVID-19 outbreak. As the autocorrelations of the conditional variances provide useful information about the heteroskedastic time series behaviour, we next introduced a suitable dissimilarity measure for the S&P 500 time series clustering. This metric is computationally efficient and simple, as it depends only on the relevant autocorrelation lags and enables the comparison of time series of possible different lengths. Additionally, regarding to the forecasting errors, we calculate the Mean Absolute Percentage Error (MAPE) of the forecasts of the index values of each industry. The forecasts are calculated from 1 to 10 steps ahead, for 23 cutoffs, each one constituted by 10 business days. As the MAPE values provide useful information about the COVID-19 impact on the stability of the time-series, we next introduced a suitable dissimilarity measure for the S&P 500 time series clustering. This metric is computationally efficient and simple, as it depends only on the 1 to 10 step-ahead forecasts MAPE values for the 11 different sub-industries, and on the 23 different cutoffs considered.

The rest of the dissertation is organised as follows. In Section 2, we briefly review the recent literature on the COVID-19 pandemic effect on the U.S. stock market volatility. In Section 3, we present a theoretical framework regarding the expectations and the impact of the COVID-19 pandemic on different U.S. industries. In Section 4, we present the data used in empirical study. In Section 5, we present the clustering methodologies and algorithms for identifying similarities and dissimilarities among financial time series. Detailed results and discussion are provided in Section 6 and, finally, we summarise the main conclusions in Section 7.

2. A BRIEF REVIEW OF LITERATURE

The first COVID-19 case in the U.S. appeared on the 21st of January 2020. According to Baker et al. (2020), COVID-19 has affected the stock market more than any other pandemic. As stated by these authors, the value of equities plummeted in the U.S., and later in the middle of March 2020, the levels of volatility saw an increase without precedent when compared with other pandemics. The volatility stabilised in April, although it remained however above the pre-pandemic levels. The most plausible reasons for this are that the restrictions imposed by the government, such as travelling restrictions, stay-at home orders, closures of non-essencial business or the use of masks. The same authors also pointed out the impact of news on the changes of volatility of the stock market. At the beginning of February, when some U.S. COVID-19-related news started to emerge, some peaks in the stock market occurred. At the end of February 2020, when the governement decreed several policies as a response to the evolution of the pandemic on the U.S., the stock market also suffered a change in stock prices.

Albulescu (2021) agrees that COVID-19 increased the volatility of U.S. financial markets. He proved empirically that two factors that contributed to this increase in volatility were the new infection cases reported at the global level and in the U.S. and also the increase in the fatality ratio. Papadamou et al. (2020) found evidence that the increased in searching for COVID-19 consequences contributed to an increased level of anxiety, which has led to a stronger negative relationship between stock market returns and their implied volatility. Another factor that influenced the market volatility amid the COVID-19 pandemic which was pointed out by Engelhardt et al. (2020), is trust in the country's government. The higher the level of societal trust and the trust in the country's government, the lower the market volatility. Engelhardt et al. (2020) explain that the higher the trust in fellow citizens and in the government, the less uncertain is the stock market of that country from investors' perspectives.

Chaudhary et al. (2020) noticed that the average returns from different stock markets of 10 distinct countries were positive from January 2019 to December 2019 and then became negative after the start of the pandemic (January 2020 – June 2020). With regards the first quarter of the period of the pandemic undweer study (January 2020 – March 2020), the authors realised that the average returns of the choosen indexes were negative, however during the second quarter (April 2020 – June 2020), they became positive and went on to discovered a significant increase in the conditional volatility of the stock market during the pandemic period, using a simple GARCH(1.1) model. Nevertheless, stock returns returned to register “normal” mean values in all markets, when considering that the mean reversion property derived by the sum of the estimated ARCH and GARCH terms was lower than 1.

Concerning the impact of COVID-19-related events in other countries on the U.S. stock market, Onali (2020) also used a GARCH(1.1) model to study the impact on the U.S. stock market of COVID-19 cases and related deaths in Italy, Spain, China, US, France, Iran, and the UK. Considering data up until the 9th of April 2020, he found that changes in the number of cases and deaths in the above-mentioned countries did not have an impact on U.S. stock market returns, apart from the number of reported cases for China, even though they impacted conditional volatility positively.

To obtain a better perspective about stock market volatility during COVID-19 in certain industries, Baek et al. (2020) used a Markov-Switching AR model to confirm a COVID-19-related change in the volatility of the U.S. stock market. To capture systematic risk, these authors used CAPM time-varying betas for each industry and concluded that there was a significant increase in systematic risk for defensive industries, such as telecom and utilities, however there was a decrease in systematic risk for automobiles and business equipment. Mazur et al. (2020) researched the stock market crash in March 2020, focussing on different industries from the S&P 500. According to these authors, those industries that performed abnormally well during this month were Healthcare, Food, Natural gas and Software. On the flip side, Crude oil, Real estate, Entertainment, and Hospitality industries were extremely negatively affected by the pandemic. They concluded that those stocks associated with these adversely affected industries have more asymmetric movements and high values of volatility which is negatively correlated with stock returns.

3. THE IMPACT OF THE COVID-19 PANDEMIC ON THE U.S. STOCK MARKET

In this dissertation, we focus on the impact of COVID-19 on the volatility of the U.S. stock market, using evidence from eleven S&P 500 industries. Those industries are Airlines, Automobile Manufacturers, Apparel, Accessories & Luxury Goods, Data Processing & Outsourced Services, Diversified Banks, Hotels, Resorts & Cruise Lines, Internet & Direct Marketing Retail, Oil & Gas Exploration & Production, Movies & Entertainment, Pharmaceuticals, and Restaurants.

The Airlines industry includes some of the United States passenger air transportation companies. The term flight frequency represents the number of flights during a specific period. As mentioned by Carter et al. (2022), airline-related companies were expected to experience a decrease in their revenue as a result of the reduction in the number of passengers and the effect of stay-at-home restrictions during the COVID-19 pandemic. The authors believe that the uncertainty regarding the duration of the pademic led to a decrease of the stock price in this industry. Figure 1 describes the change of weekly flight frequency of U.S. Airlines. According to Statista (2021a), the airline industry was thriving in the U.S. before the pandemic, which was reflected in a projected annual growth in revenue of 4% from 2020 to 2040 in the United States. However, as reported in the Bureau of Transportation Statistics (2021), the effect of lockdowns and the various restrictions imposed in some places due to the COVID-19 pandemic caused a decrease in the net profit of U.S. Airlines of $49.8 billion when compared to the net profit of 2019. The total loss for U.S. Airlines in 2020 was $35 billion.

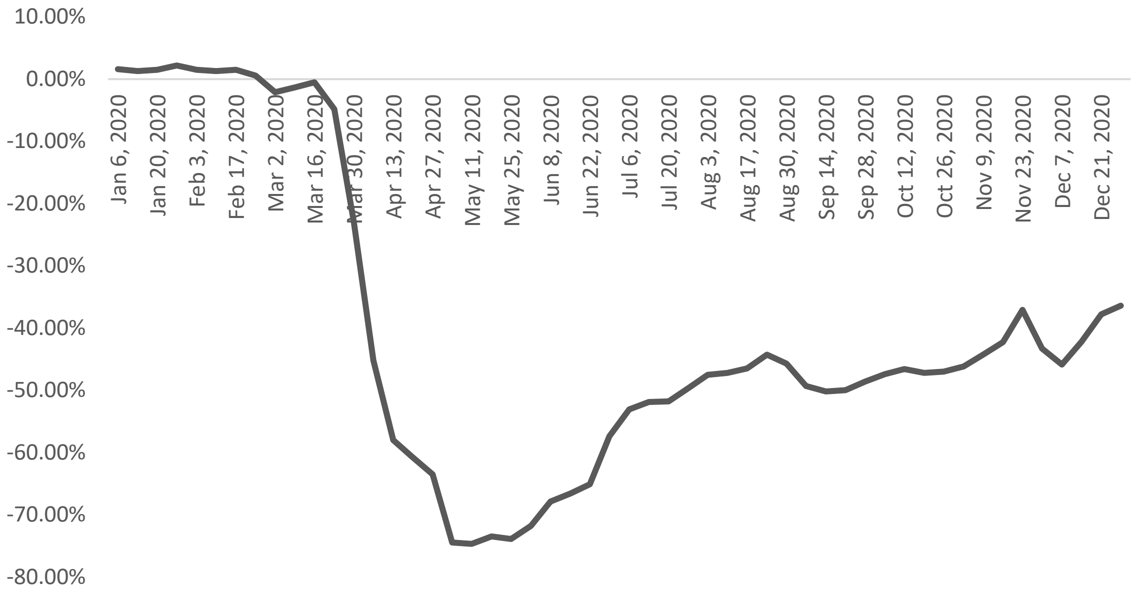

FIGURE 1 - Change of weekly flight frequency of U.S. Airlines. Source: Statista (2021a)

As Figure 1 reveals, the change of weekly flight frequency of U.S. Airlines was negative from late February up until the end of December, which explains the total losses that this industry suffered. For this reason, it would be relevant to study how COVID-19 impacted the stock market returns of those airlines tracked by the S&P 500.

The Apparel, Accessories & Luxury Goods industry is represented by the producers of designer handbags, wallets, luggage, jewellery, and watches. This industry was expected to suffer a negative shock following the COVID-19 pandemic, as stated by Achille & Zipser (2020). A range of factors support this statement. Firstly, the North American luxury sector was already struggling before the pandemic started, with the pandemic worsening sales. In turn, wholesalers were likely to implement aggressive discount policies in order to survive, which would hurt those companies that did not have a concession model. Another reason presented by Achille & Zipser (2020) is that 20% to 30% of luxury sector revenues are raised by consumers from countries other than the U.S., where revenues are expected to decrease owing to travel restrictions.

The Automobiles Manufacturers industry includes companies that produce light trucks and passenger automobiles, but does not include companies which produce motorcycles, three-wheelers, and heavy-duty trucks. According to Deloitte (2020a), COVID-19 had a negative impact on the Automotive industry, as manifested by the closure of many assembly plants in the U.S., which caused distress for the global supply base.

The Diversified Banks industry is constituted by U.S. banks with revenues primarily originating from conventional banking operations which have business activity in retail banking that provide a diverse range of financial services. Deloitte (2020b) mentioned there are some potential long-term COVID-19 impacts on banks, such as a continued reduction of interest rates, business activities, and also a high number of non-performing loans. This negative shock on the Banking industry will be analysed using this industry as a proxy of the U.S. investment and retail banks.

The Hotels, Resorts, and Cruise Lines industry represents the owners and operators of hotels, resorts, and cruise-ships. Krishnan et at. (2020) pointed out that this industry would be one of the most negatively affected by the pandemic, mainly due to the impact of changes in travel restrictions, consumer sentiment, and willingness to travel.

The Internet & Direct Marketing Retail industry tracks those companies that provide retail services through the internet. The data available from Statista (2022a) shows evidence that the percentage of e-commerce in terms of total retail sales was the highest ever recorded between March and April 2020. Before the onset of COVID-19, e-commerce as a share of total retail sales was 11% and at the start of the pandemic it was actually 22% between March and April 2020, due to certain COVID-19 pandemic restrictions and lockdowns and it is expected that the customers will increased online shopping to the detriment of in-person shopping. For the same reasons, the Data Processing & Outsourced Services industry was also included in this study, which englobes the set of companies that provide commercial electronic data processing and/or business process outsourcing services. With an increase in online shopping and online payments, the impact of COVID-19 is expected to be positive in this industry and this will be assessed further on below.

The Movie & Entertainment industry is composed of companies that operate in the movie industry, selling and distributing entertainment content, such as movies, TV shows, music, and sports. This industry was chosen to check whether the COVID-19 pandemic impact was either negative or positive in these companies. Despite the closure of many U.S. theatres, some of the companies in that industry possess streaming platforms, such as Netflix and Disney+, which saw the number of their subscribers increasing at the start of the pandemic. According to data from the Statista website (2022b), the number of Netflix's paid subscribers worldwide during the 1st Quarter of 2020 was 182.86 million, which means that it gained 15.77 million new subscribers compared to the 4th Quarter of 2019. Figure 2 shows the increase in Netflix spectator numbers from the 1st Quarter of 2016 to the 4th Quarter of 2020. Figure 3 describes the number of Disney+ subscribers from the 1st Quarter of 2020 until the 4th Quarter of the same year, according to data from the Statista website (2022b).

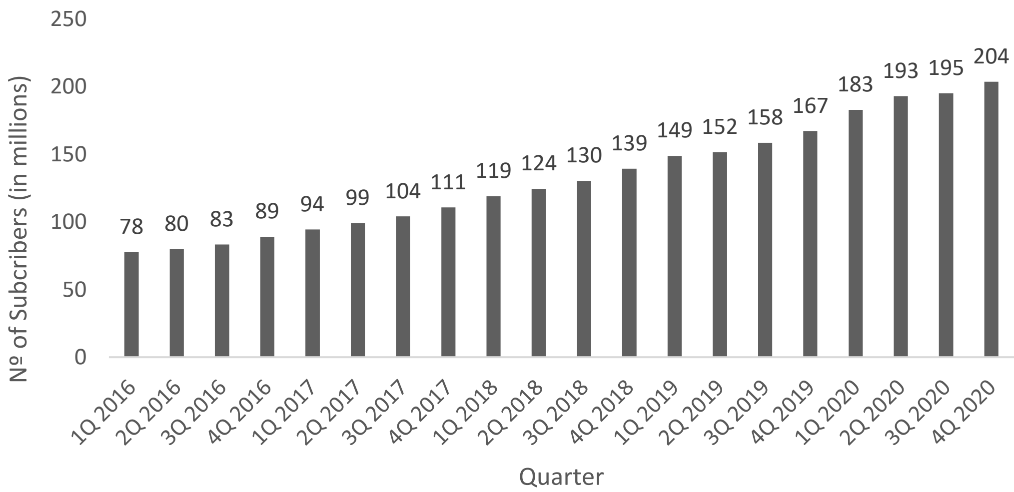

FIGURE 2 - Netflix spectator numbers from 2016 to 2020. Source: Statista (2022b)

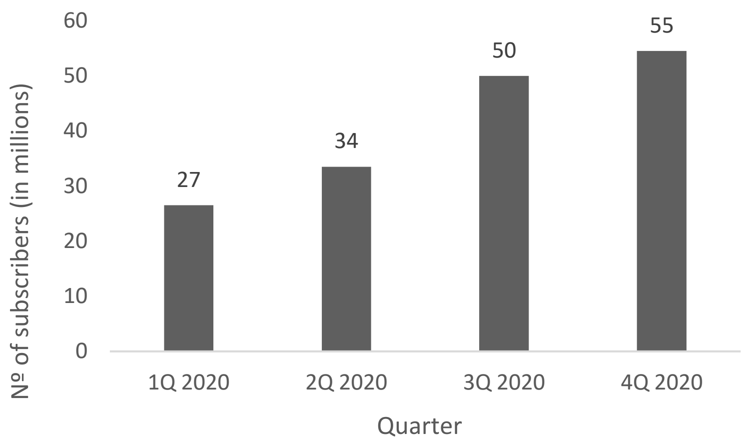

FIGURE 3 - Disney+ subscribers during 2020. Source: Statista (2021b)

During the first year of the pandemic in 2020, Netflix increased its number of subscribers from approximately 183 million to 204 million. As can be seen in both figures, these two streaming platforms increased the number of subscribers during the pandemic. On the other hand, Disney+, which was launched by The Walt Disney Company on the 19th of November 2019, increased its worldwide subscribers from 27 million to 55 million over the same period. These events might well have affected the stock returns of the Movie & Entertainment industry, which will be studied in more detail further.

The Oil & Gas & Exploration & Production industry englobes some of the companies engaged in the exploration and production of oil and gas. As mentioned by Deloitte (2020c), an imbalance related with the supply/demand of oil due to the existence of slow-downs and travel restrictions related with the COVID-19 pandemic we will attempt to measure the impact the spread of COVID-19 on this industry.

The Pharmaceuticals industry represents the set of companies involved in the research and production of pharmaceuticals, including veterinary drugs. Certain safety and health products for preventing COVID-19 infection were developed and produced by pharmaceutical companies. Mittal & Sharma (2021) studied the impact of COVID-19 on the stock returns of the Indian pharmaceutical sector, comparing them for the period before the 12th of March, when the first pandemic fatality died from the virus, and from then onwards. They concluded that the pharmaceuticals sector returns were positive and statistically significant between the 1st of February 2020 and the 24th of April 2020. Seemingly, contrarily to other sectoral indices, the Pharmaceutical industry's appears to have thrived in India during the beginning of the pandemic. In this study we also investigate whether the same happened in the U.S.

The Restaurants industry is constituted by owners and operators of restaurants, fast-food enterprises, bars, and take-out facilities. According to Ding at et al. (2021), an increase in leverage ratio also suggests a rise in the level of risk for businesses, which in turn leads to a decrease in stock prices due to investor anxiety. However, Song (2021) contradicts this perspective, stating that the increase in leverage in restaurant companies during COVID-19 positively affected stock returns, as shareholders could be compensated with extra value due to tax benefits arising from leverage.

4. DATA

The objective of this section is to provide a better understanding of how the companies tracked by Standard & Poor's 500 (S&P 500) are split into different sectors and industry groups, according to the Global Industry Classification Standard (GICS). The GICS is a method created by Standard & Poors and Morgan Stanley Capital International (MSCI) to attribute an economic sector and industry group to companies, based on their primary business activity.

This classification standard is an important tool for many financial institutions or investors, as it enables one to know who are the main competitors for a specific company within the same sector and industry. In the scope of this study, this classification standard system is used to track the time-varying volatility of 11 S&P 500 industries, before and during the spread of the COVID-19 pandemic. This analysis will focus on those industries and also those companies included in each of these industries, which is summarised in Table I. All the data for S&P 500 industries index levels was obtained from MarketWatch, which is a subsidiary of Dow Jones & Company which provides financial content regarding stock market data.

The full period for which this study will focus on is from 05/01/2016 until 11/12/2020. The time set that represents the period before the pandemic is from 05/01/2016 until 17/01/2020 and the time set that embodies the pandemic period starts on 21/01/2020, which is the date of the detection of the first U.S. citizen infected with coronavirus, up until 11/12/2020, which was the day of the vaccination of the first American citizen. The time series plots of industries index levels are shown in Figure 4.

TABLE I - GICS classification system: Sector, industry, sub-industry, and selected companies

Sector	Industry / Sub-industry	Companies	
Communication Services	Entertainment / Movies & Entertainment	Fox Corporation (class A) (FOXA) Fox Corporation (class B) (FOX) Live Nation Entertainment (LYV)	Netflix (NFLX) The Walt Disney Company (DIS) ViacomCBS (VIAC)
Consumer Discretionary	Internet & Direct Marketing Retail/ Internet & Direct Marketing Retail	Amazon (AMZN) Booking Holdings (BKNG) eBay (EBAY)	Etsy (ETSY) Expedia Group (EXPE)
	Automobiles / Automobile Manufacturers	Ford (F) GM (GM)	Tesla (TSLA)
	Textiles, Apparel & Luxury Goods / Apparel, Accessories & Luxury Goods	Nike (Nike) PVH (PVH) Ralph Lauren Corporation (RL) Tapestry (TPR)	Under Armour (class A) (UAA) Under Armour (class C) (UA) VF Corporation (VFC)
	Hotels, Restaurants & Leisure / Hotels, Resorts & Cruise Lines	Carnival Corporation (CCL) Hilton Worldwide (HLT) Marriott International (MAR)	Norwegian Cruise Line Holdings (NCLH) Royal Caribbean Group (RCL)
	Hotels, Restaurants & Leisure / Restaurants	Chipotle Mexican Grill (CMG) Darden (DRI)	Domino's (DPZ) McDonald's (MCD)
Healthcare	Pharmaceuticals / Pharmaceuticals	Johnson & Johnson (JNJ) Pfizer (PFE)	AbbVie (ABBV) Viatris (VTRS)
Industrials	Airlines / Airlines	Alaska Air Group (ALK) American Airlines Group (AAL) Delta Air Lines (DAL)	Southwest Airlines (LUC) United Airlines (UAL)
Energy	Oil, Gas & Consumable Fuels / Oil & Gas Exploration & Production	Baker Hughes (BKR) APA Corporation (APA) Halliburton (HAL) Schlumberger (SLB) ConocoPhillips (COP) Coterra (CTRA)	Devon Energy (DVN) Diamondback (DVN) EOG Resources (EOG) Marathon Oil (MRO) Occidental Petroleum (OXY) Pioneer Natural Resources (PXD)
Financial	Banks/Diversified Banks	Bank of America (BAC) Citigroup (C) Comerica (CMA) JPMorgan Chase (JPM)	U.S. Bancorp (USB) Wells Fargo (WFC) Eastman (EMN)
Information Technology	IT Services/Data Processing & Outsourced Services	ADP (ADP) (BR) FIS (FIS) Fiserv (FISV) Fleetcor (FLT) Global Payments (GPN)	Jack Henry & Associates (JKHY) Mastercard (MA) Paychex (PAYX) PayPal (PYPL) Visa (V)

Note: The sub-industries considered to study the impact of the COVID-19 on their stock prices are presented in column 2. The last column contains a brief description of each sub-industry.

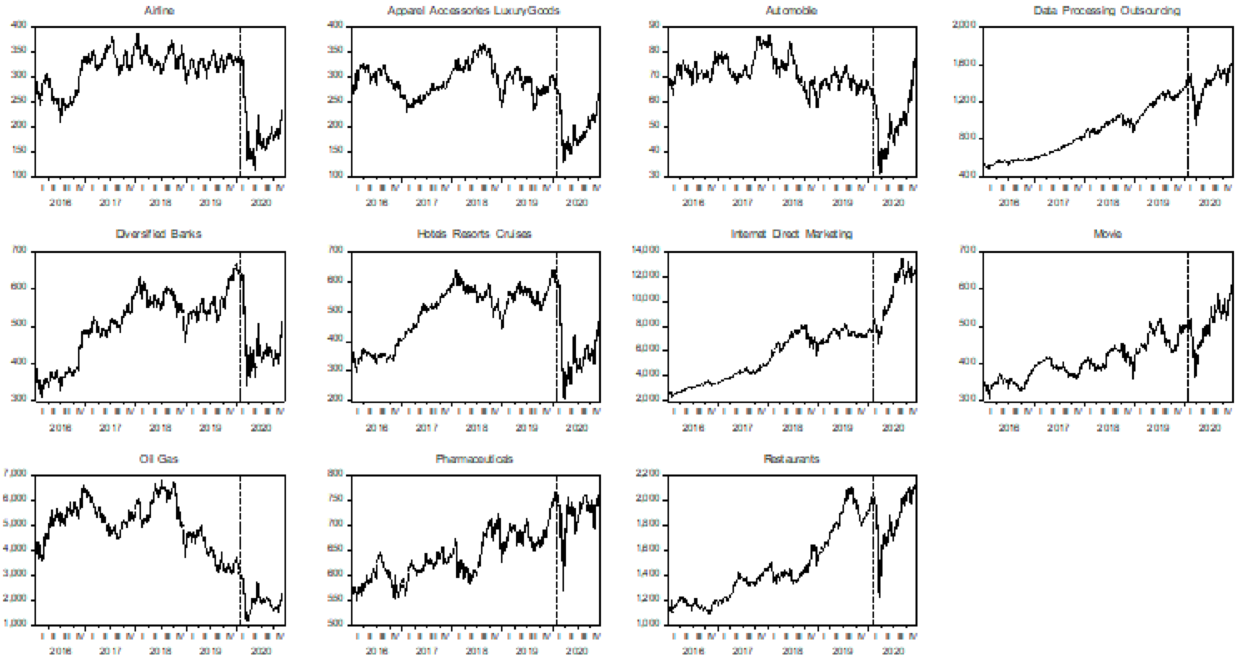

FIGURE 4 - Time series plots of the daily S&P 500 industry indices. The vertical dotted line separates the periods before and after the spread of COVID-19 in the United States

In order to calculate the forecast errors, our data had to be divided in 23 different cutoffs throughout 2020, each one with ten days. The constitution of each cutoff is presented in Table II.

TABLE II - Cutoff constitution: start and ending date, and number of business days included

Cutoff (Ci)	**Start Date**	**End Date**	**Number of Business Days**
C1	21/01/2020	03/02/2020	10
C2	04/02/2020	18/02/2020	10
C3	19/02/2020	03/03/2020	10
C4	04/03/2020	17/03/2020	10
C5	18/03/2020	31/03/2020	10
C6	01/04/2020	15/04/2020	10
C7	16/04/2020	29/04/2020	10
C8	30/04/2020	13/05/2020	10
C9	14/05/2020	28/05/2020	10
C10	29/05/2020	11/06/2020	10
C11	12/06/2020	25/06/2020	10
C12	26/06/2020	10/07/2020	10
C13	13/07/2020	24/07/2020	10
C14	27/07/2020	07/08/2020	10

C15	10/08/2020	21/08/2020	10
C16	24/08/2020	04/09/2020	10
C17	08/09/2020	21/09/2020	10
C18	22/09/2020	05/10/2020	10
C19	06/10/2020	19/10/2020	10
C20	20/10/2020	02/11/2020	10
C21	03/11/2020	16/11/2020	10
C22	17/11/2020	01/12/2020	10
C23	02/12/2020	11/12/2020	8

Note: Contrarily to all the other cutoffs, the last one just takes into consideration 8 business day

5. METHODOLOGY

In this section we introduce the clustering approach in order to identify similarities and dissimilarities among time series with heteroskedastic dependence structure. This includes the choice of a distance (or dissimilarity) measure between stock market industries and the choice of a clustering method to group data into distinct clusters.

Let P_t denote the price of an asset at time *t*. The continuously compounded return (or log return) from $t - 1$ to t is defined as

(1) $$r_t = \log\left(P_t / P_{t-1}\right).$$

The squared return r_t^2 (or the absolute return $|r_t|$) is usually considered as a proxy of volatility. Many empirical studies (e.g., Ding et al., 1993, Granger and Ding, 1996 and Engle and Patton, 2001) have noticed slowly decaying autocorrelations for squared (or absolute) financial returns. The autocorrelation function of a time series x_t at lag k is defined by

(2) $$\rho_k = \frac{Cov(x_t, x_{t+k})}{\sqrt{[Var(x_t)][Var(x_{t+k})]}} = \frac{\gamma_k}{\gamma_0}, k = 0,1,2, \ldots,$$

where

(3) $$\gamma_k = Cov(x_t, x_{t+k}) = E[(x_t - \mu)(x_{t+k} - \mu)]$$

is the autocovariance functions. The autocorrelation function is an even function, $\rho_k = \rho_{-k}$, and has its maximum magnitude at lag $k = 0$, $|\rho_k| \leq \rho_0 = 1$.

In this study, the presence of conditional volatility dependence in S&P500 industry indices is judged by the autocorrelations of the fitted GARCH (generalized autoregressive heteroskedasticity) variance. This conditional variance (or volatility) is estimated using a parsimonious threshold GARCH (or TGARCH) model (see Glosten et al., 1993; Zakoian, 1994) to allow for asymmetric shocks to volatility. The simple TGARCH (1.1) model assumes the form

(4) $$\varepsilon_t = \sigma_t z_t,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma d_{t-1} \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where $\{z_t\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance $d_t = 1$ if ε_t is negative, and $d_t = 0$ otherwise. The volatility may be either diminished ($\gamma < 0$), rise ($\gamma > 0$), or not be affected ($\gamma = 0$) by negative shocks or "bad news" ($\varepsilon_{t-1} < 0$). "Good news" has an impact of α while "bad news" has an impact of $\alpha + \gamma$. The persistence of shocks to volatility is given by $\alpha + \beta + \gamma/2$.

Cluster analysis of financial time series is concerned with exploring data sets (observation-based, feature-based, and model-based methods) to discover natural groups or clusters of financial assets to ensure that they are as similar as possible within the same group and are different from assets in other groups. The choice of a relevant measure of similarity between each pair of time series and its features is key for cluster analysis. In Finance, it seems reasonable to apply classification and clustering methods that best capture the specific features of financial time series, such as autocorrelations, nonlinearities, volatility, and other stylised facts.

For this purpose, we introduce a new distance measure between two financial time series x and y, based on conditional volatility dependence, which is defined by

$$(5) \qquad d_{ACFCV}(x,y) = \sqrt{\left(\hat{\rho}_{\hat{\sigma}^2_{t,x}} - \hat{\rho}_{\hat{\sigma}^2_{t,y}}\right)' \Omega \left(\hat{\rho}_{\hat{\sigma}^2_{t,x}} - \hat{\rho}_{\hat{\sigma}^2_{t,y}}\right)},$$

where $\hat{\rho}_{\hat{\sigma}^2_{t,x}}$ and $\hat{\rho}_{\hat{\sigma}^2_{t,y}}$ are the sample autocorrelation functions of the fitted conditional variance TGARCH models to assets x and y, respectively, and Ω is a weighting matrix. If $\Omega = I$, we obtain the Euclidean distance between the autocorrelations of x and y. If $\Omega = M^{-1}$, we get the Mahalanobis distance between the autocorrelations, where M is the sample covariance matrix of the autocorrelation coefficients. The metric $d_{ACFCV}(x,y)$ satisfies the minimal properties we could expect of a distance: non-negativity, $d(x,y) \geq 0$; symmetry, $d(x,y) = d(y,x)$; and triangle inequality, $d(x,z) \leq d(x,y) + d(x,z)$.

A wide range of clustering algorithms can be applied for distances among these vectors of sample autocorrelations. The most used procedures include hierarchical and non-hierarchical clustering algorithms. In our study, we implement hierarchical complete linkage and K-means methods. In the former, one chooses the pair of clusters to merge at each step that are separated by the shortest distance, and in the latter, one determines a preliminary set of k clusters, move each time series to the cluster whose centroid is closest in Euclidean distance, recalculate the cluster centroid and repeat the reassignment procedure until no time series is reassigned. For a detailed discussion of these clustering methods, see, for example, Everitt et al. (2001).

Then, by comparing and clustering financial time series according to their forecast error structure. We use fitted TGARCH models to compute multi-step-ahead forecasts of stock market indices for different cutoff points. The accuracy of these forecasts is measured using the Mean Absolute Percentage Error (MAPE), which is defined by

(6) $$MAPE = \frac{1}{n-m} \sum_{t=m+1}^{n} \frac{|Y_t - F_t|}{Y_t} \times 100,$$

where Y_t is the actual value for time period t and F_t is the forecasted value for the same period. This statistic is used to evaluate the out-of-sample forecast accuracy using a training sample of observations of size m < n (where n is the sample size) to estimate the model, and next computing recursively the one-step ahead forecasts for time periods $m+1, m+2, \ldots$ by increasing the training sample by one. For k-step ahead forecasts, we begin at the start of the training sample and compute the forecast errors for time periods $t = m+k, m+k+1, \ldots$ using the same recursive procedure.

It is well-known that cluster analysis is a commonly-used method for finding similarities and dissimilarities between pairs of time series in a data set. It comprises the following steps: first, we collect data of a set of time series and measure their features, where for Finance, the attributes that best capture the specific features of financial time include autocorrelations, periodogram ordinates, nonlinearities, volatilities, and other stylised facts; in the second step we need to define and compute the distances (or similarities) between each pair of time series and its features (or standardised features), where the Euclidean (or standardised Euclidean) distance is the most-used resemblance measure in cluster analysis, which measures the distance between two objects or time series in the two-dimensional space formed by their features; finally, the third step requires choosing a clustering method for creating the clusters, where the choice depends on whether a dissimilarity measure or a similarity measure is used, bearing in mind that there are different types of clustering methods, including partitioning methods, hierarchical clustering, fuzzy clustering, density-based clustering, feature-based clustering and model-based clustering.

Then, we introduce a new model-based metric for clustering of financial time series which uses the forecast errors of fitted volatility models to compute the distance between every pair of time series. Similar definitions and distributional results apply to time series $y_t, t = 1,2, \ldots, n_y$. $MAPE_x$ and $MAPE_y$ are the vectors of mean absolute percentage errors of the multi-step ahead TGARCH model forecasts for assets x and y, respectively. The distance between the forecast errors of x and y is defined by

$$(7) \qquad d_{MAPE}(x, y) = \sqrt{\left(MAPE_x - MAPE_y\right)' \Omega \left(MAPE_x - MAPE_y\right)},$$

where Ω is some matrix of weights. If Ω=I, when I is an identity matrix we obtain the Euclidean distance between the multi-step-ahead forecast errors of the time series of x and y. If Ω=V^{-1}, where V is a diagonal matrix with the variances of the forecast errors, we obtain the standardised Euclidean distance between the forecast errors of the time series of x and y. If Ω=C^{-1}, where C is the sample covariance matrix of the forecast errors, we obtain the Mahalanobis distance between x and y. It is straightforward to show that this metric satisfies the usual properties we could expect of a distance, i.e., non-negativity, symmetry, and triangle inequality. A wide range of clustering algorithms can be applied for distances among these vectors of MAPE. In our study, we investigate the affinity of time series models using the hierarchical complete and single linkage dendrograms associated with the underlying standardised forecast errors. For a detailed discussion of these clustering methods, see, for example, Everitt et al. (2001).

6. RESULTS AND DISCUSSION

Tables III and IV present standard unconditional and conditional statistics of each industry before and during the COVID-19 pandemic. The standard unconditional descriptive statistics of the S&P 500 log returns include the annualised mean, the annualised standard deviation, skewness, and the excess kurtosis of returns. The mean is computed as the annualised average log return. The standard deviation, or unconditional volatility, is a measure of dispersion in the return series, and is usually considered to be a proxy for asset risk. Skewness is the coefficient of the asymmetry of the distribution of the return series. Kurtosis measures the "fatness" of the tails of the return distribution, where a distribution with positive excess kurtosis has heavy tails and a distribution with negative excess kurtosis has short tails. Conditional statistics include the TGARCH models estimates that are described in Section 5.

Only three (Internet & Direct Marketing Retail, Movie & Entertainment, Oil & Gas Exploration & Production) of the 11 industries achieved an increase in their mean log returns from the pre-COVID-19 period to the post-COVID-19 period. As expected, unconditional volatility for all the industries increased from the period before the pandemic to the period when the pandemic starts. Three industry indices (Airlines, Hotels and Resorts & Cruise Lines, and Restaurants) experienced increases in unconditional volatility of more than 200% and four industry indices (Apparel, Accessories & Luxury Goods, Automobile Manufacturers, Data Processing & Outsourcing Services, and Diversifying Banks) experienced increases greater than 150%. All industries (except Oil & Gas Exploration & Production in pre-COVID-19 and Pharmaceuticals in post-COVID) exhibit negative skewness, indicating that the distributions of returns have long left tails. Restaurants, Oil & Gas Exploration & Production, and Data Processing & Outsourcing Services exhibit the highest kurtosis coefficients (16.3, 11.7, and 10.1, respectively) during the spread of COVID-19. In both periods, the volatility tends to be higher in the presence of negative shocks and the estimates of the persistence of shocks to volatility are very close to one. Only the Automobile Manufacturers industry exhibits a negative correlation between volatility and negative shocks in the period before COVID-19. During the spread of COVID-19, the sum of the TGARCH terms exceeds 1 for

Diversified Banks and Pharmaceuticals industries, which means that the volatility increases exponentially over time.

TABLE III - Summary unconditional (mean, standard deviation, skewness, kurtosis) and conditional (arch, garch, leverage, persistence) statistics for the S&P500 indices before the COVID-19 pandemic

	Log returns				Fitted TGARCH(1,1) model			
Sector	**mean**	**st. dev.**	**skew**	**kurt**	**arch**	**garch**	**lever.**	**persist.**
Airlines	5.06%	25.38%	-0.385	4.954	-0.011	0.920	0.069	0.944
Apparel, Ac. & Luxury Goods	0.22%	23.15%	-0.549	5.286	0.049	0.821	0.093	0.916
Automobile Manufacturers	-3.92%	23.40%	-0.066	5.146	0.145	0.532	-0.026	0.664
Data Proc. & Outsourc. Services	24.59%	17.92%	-0.339	6.457	0.044	0.767	0.223	0.923
Diversifying Banks	12.37%	20.67%	-0.255	5.947	0.022	0.807	0.169	0.913
Hotels, Resorts & Cruise Lines	13.56%	20.77%	-0.788	6.769	0.005	0.896	0.095	0.949
Internet & Direct Mkt. Retail	24.09%	25.01%	-0.190	7.093	0.045	0.812	0.163	0.938
Movie & Entertainment	9.16%	18.03%	-0.278	6.292	0.046	0.908	0.055	0.982
Oil & Gas Exploration & Prod.	-9.06%	34.34%	0.114	4.770	-0.019	0.985	0.048	0.990
Pharmaceuticals	6.45%	13.99%	-0.487	6.754	0.004	0.911	0.097	0.964
Restaurants	14.19%	14.04%	-0.448	5.879	0.073	0.816	0.075	0.927

TABLE IV - Summary unconditional (mean, standard deviation, skewness, kurtosis) and conditional (arch, garch, leverage, persistence) statistics for the S&P500 indices during the COVID-19 pandemic

	Log returns				Fitted TGARCH(1,1) model			
Sector	**mean**	**st. dev.**	**skew**	**kurt**	**arch**	**garch**	**lever.**	**persist.**
Airlines	-37.52%	79.96%	-0.326	6.674	0.000	0.831	0.144	0.903
Apparel, Ac. & Luxury Goods	-10.78%	62.38%	-0.235	4.980	0.008	0.853	0.226	0.974
Automobile Manufacturers	11.08%	59.16%	-0.088	7.964	0.041	0.838	0.171	0.965
Data Proc. & Outsourc. Services	9.23%	44.80%	-0.185	10.134	0.023	0.777	0.307	0.954
Diversifying Banks	-25.63%	59.75%	-0.077	7.338	0.065	0.799	0.283	1.006
Hotels, Resorts & Cruise Lines	-33.62%	74.80%	-0.036	5.298	-0.012	0.919	0.161	0.988
Internet & Direct Mkt. Retail	51.54%	38.39%	-0.145	4.035	0.200	0.649	0.057	0.878
Movie & Entertainment	31.50%	39.92%	-0.540	7.145	-0.006	0.796	0.213	0.896
Oil & Gas Exploration & Prod.	-32.13%	79.29%	-0.831	11.718	0.066	0.804	0.154	0.946
Pharmaceuticals	2.07%	29.48%	0.096	7.140	-0.060	0.858	0.421	1.012
Restaurants	5.46%	42.31%	-0.179	16.310	0.096	0.704	0.323	0.962

Figure 5 plots the industries' conditional volatilities for the period of 04/01/2016-11/12/2020. The vertical red line separates the period before and after the start of the pandemic in the U.S. We can notice a significant increase in volatility during the start of the pandemic in all industries, except for the Internet & Direct Marketing Retail industry.

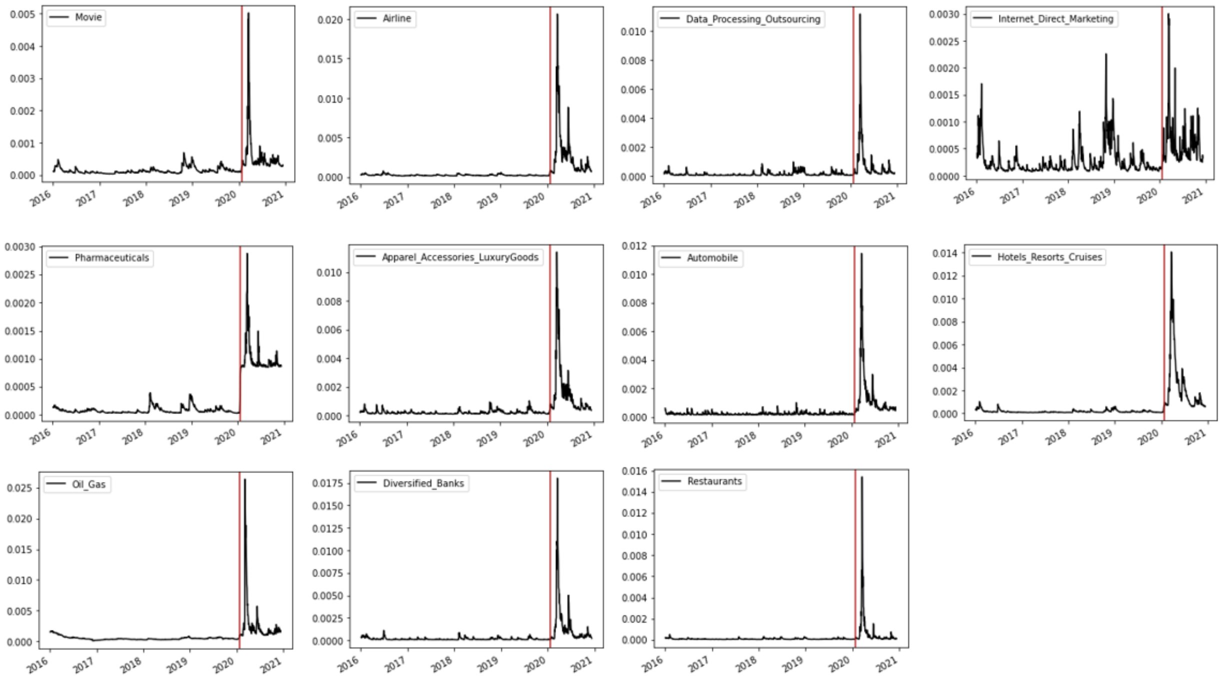

FIGURE 5 - Time series plots of the conditional volatilities of the S&P 500 industry indices

Regarding the Movie & Entertainment industry, the index levels for this industry suffered a sharp decrease during the spread of COVID-19, leading to an increase in conditional volatility. This occurred because, with the exception of Netflix, all the other companies inside the Movie & Entertainment industry experienced a decrease in their prices during 2020. Nevertheless, mainly due to the existence of streaming platforms, such as Netflix and Disney+, this industry index level quickly recovered. The behaviour of the pharmaceuticals industry is similar. At the beginning of the pandemic most of the companies' stock prices steeply decreased, contributing to an increase in conditional volatility during the same period and then quickly recovered, which could be justified by a surge in demand for pharmaceutical COVID-19 related products, such as masks and hand sanitisers. Another factor which could have contributed to this increase is the news regarding vaccines disseminated by some companies, such as Pfizer and Johnson & Johnson.

The Airlines industry's conditional volatility strongly increased during the spread of COVID-19, as expected, due to the sharp decrease of the weekly flight frequency of U.S. Airlines (in late February). The conditional volatility of the Apparel, Accessories & Luxury goods industry increased during the COVID-19 period, as a result of the aggressive discounting policies implemented by wholesalers and also the travel restrictions that prevented some consumers located out of the U.S. to reach this market. As expected, volatility also increased in the Hotels, Resorts and Cruise Lines industry, due to a decrease in willingness to travel and related restrictions, as well as consumer sentiment. The growth in volatility of the Oil & Gas & Exploration & Production industry during the COVID-19 spread most likely occurred for the above-mentioned reasons. As a result of the decrease in interest rates and the large amount of non-performing loans during the COVID-19 period, the Diversified Banks industry was expected to be more volatile than before the pandemic, which is exactly what ended up happening. The Restaurants industry also suffered from high volatility, which was expected, due to sanitary restrictions. The Data Processing & Outsourced Services industry experienced a brief decrease in the index level at the beginning of the pandemic, which resulted in an increase of the conditional volatility, contrarily to what was expected. On the other hand, the Internet & Direct Marketing Retail industry maintained similar volatility levels as before, as according to our expectations. Negative shocks usually have a stronger impact on volatility, and therefore its absence could justify why volatility did not change as much as it did in other industries.

To study the correlation between these industries in more detail, the cluster heatmaps of the conditional volatility correlations were computed for both samples, which are presented in Figure 6. The cluster heatmaps show evidence of the change in correlation between the volatility of the chosen industries. The clearest change in the correlations was between the Automobile Manufacturers and Oil & Gas Exploration and Production industries, which increased from 0.18 to 0.61. The correlation between the volatilities of the Movie & Entertainment and the Diversified Banks industries increases from 0.5 to 0.94, becoming almost perfectly correlated. In general, all the volatility correlations increased from one period to the other. The industries more affected by this increase were the Movie & Entertainment, Diversified Banks, Data Processing & Outsourcing, Apparel, Accessories & Luxury Goods and the Automobiles Manufacturers industries.

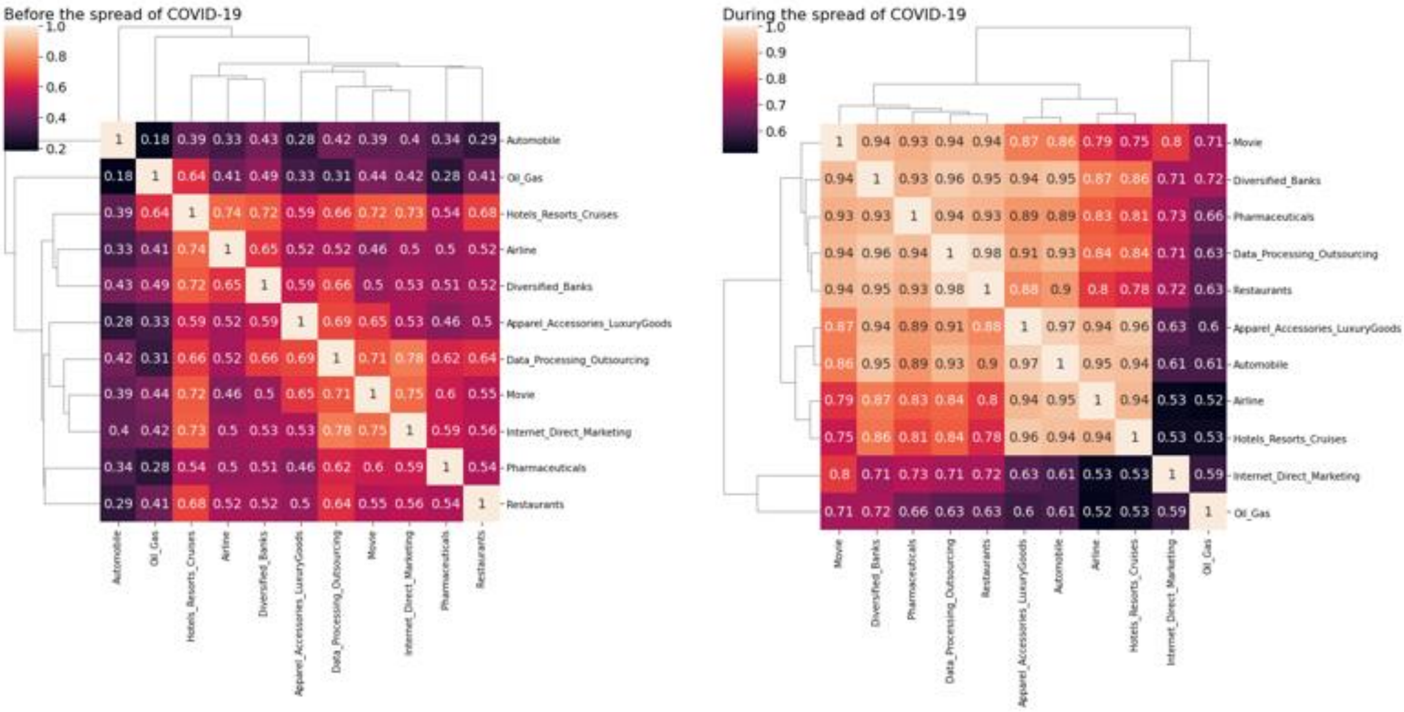

FIGURE 6 - Cluster heatmap of the conditional volatility correlations for the daily S&P 500 industry indices before (left) and during (right) the spread of COVID-19

In our cluster analysis study, we first fit a principal component model to the autocorrelations of the estimated volatilities. The principal component loadings for the S&P500 industries before and during the spread of COVID-19 are shown in Figure 7. Both before and during the COVID-19 pandemic, the most relevant autocorrelations in Principal Component 1 are autocorrelations 6 to 13 (dependence on volatility of about 2 to 2.5 weeks). Regarding Principal Component 2, before the COVID-19, the most significant autocorrelations are those from Lags 2 to 8 (short-term dependence) and from Lags 18 to 21 (about 4 weeks), while during the spread of COVID-19 they are the autocorrelation from Lag 3 to 7 (about 0.5 to 1.5 weeks) and from Lag 13 to 14 (about 3 weeks). Finally, with regards Principal Component 3, before the spread of COVID-19 the most relevant autocorrelations values are the ones from Lag 1 to 3, Lag 6 to 8, Lag 13 to

16 and Lag 21, while during the pandemic, for the same principal component, these autocorrelations are more significant from Lag 8 to 11 and from Lag 18 to 21. During the COVID-19 period, in Principal Component 3, the autocorrelations from Lag 18 to 21 assume negative values contrary to what occurred during the period before the start of the COVID-19 pandemic.

We compute the proportion of total variance explained by each component and the cumulative sum explained by the variance of the principal components. We retain the first components that explain more than 90% of the total variance in the data and then we use these components to fit both hierarchical (complete linkage) and non-hierarchical (K-means) clustering models. As the number of trading days in a month for the U.S. market is 21 days on average, we calculated the Euclidean distance by only using the first 21 sample autocorrelations of the fitted TGARCH variance, to ascertain whether volatility dependence similarities and dissimilarities among S&P500 industries exist.

Table V shows the Silhouette and Davis-Bouldin validity criteria for determining the number of clusters. Both metrics suggest an optimal number of clusters of 4. Table VI presents the K-means cluster allocation of S&P 500 industries before and during the spread of COVID-19. Similar results are obtained using the hierarchical complete linkage algorithm, which minimises the maximum distance between industries in the same cluster, as shown in the dendrograms in Figure 8. The dendrograms show how clusters are formed at each stage of the procedure. At the bottom of the dendrogram, each object is considered to be its own cluster and these objects continue to join together as they head upwards. At the top, all the objects are grouped into a single cluster. In hierarchical clustering, partitions are obtained by cutting off the dendrogram at an arbitrary point. The choice of the appropriate number of clusters in the dendrogram is sometimes subjective and depends on the expert judgment of the researchers. Figure 9 displays the two-dimensional score plots of S&P500 industries which are obtained by the two principal components that explain the largest proportion of the total variance. Two clusters of industries were identified before the COVID-19 pandemic. The first cluster is composed of the Consumer Durables & Apparels, Financials, and Information Technology industries. The second cluster is composed of the Healthcare, Consumer Services, and Consumer Discretionary industries. The Automobile and Energy industries are separated from each other in two isolated clusters. During the spread of COVID-19, the Internet &

Direct Marketing Retail industry clearly distanced itself from the other industries, especially from those industries most affected by the pandemic (Hotels, Apparel & Luxury Goods, Automobile and Airline).

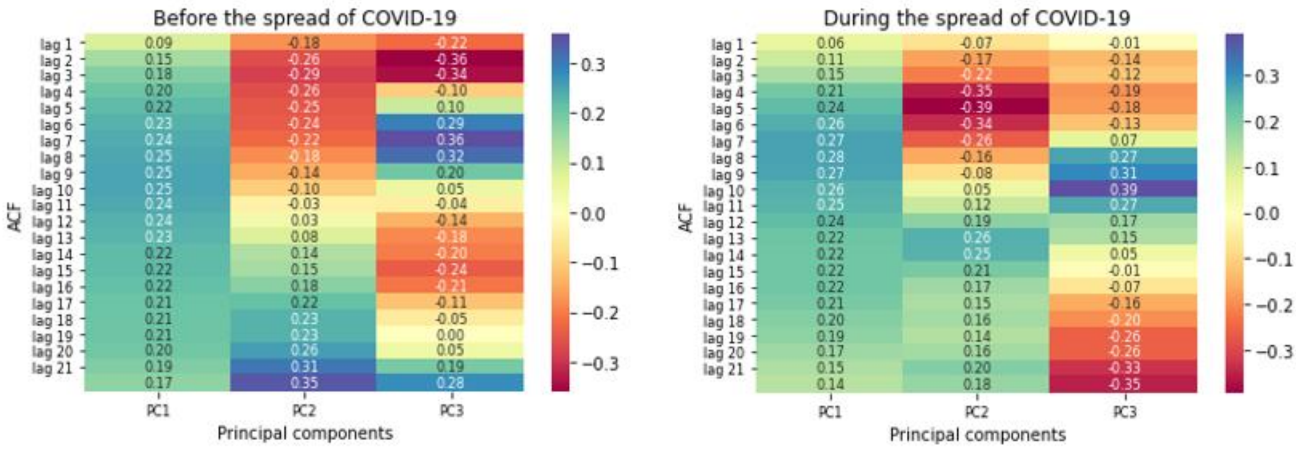

FIGURE 7 - PCA loadings for the autocorrelations of the estimated conditional variances on daily S&P 500 industry indices before (left) and during (right) the spread of COVID-19

TABLE V - Silhouette and Davis-Bouldin validity criteria for determining the number of clusters

Criteria	Before the spread of COVID-19	During the spread of COVID-19
Silhouette scores	0.526 (K=2) 0.509 (K=3) **0.537 (K=4)** 0.421 (K=5)	0.459 (K=2) 0.461 (K=3) **0.464 (K=4)** 0.356 (K=5)
Davis-Bouldin scores	0.638 (K=2) 0.425 (K=3) **0.273 (K=4)** 0.274 (K=5)	0.666 (K=2) 0.585 (K=3) **0.423 (K=4)** 0.445 (K=5)

TABLE VI - K-means cluster solution for the daily S&P 500 industry indices

Cluster	Before the spread of COVID-19	During the spread of COVID-19
1	Data Proc. & Outsourcing Services; Apparel, Accessories & Luxury Goods; Diversifying Banks; Restaurants	Airlines; Data Proc. & Outsourcing Services; Pharmaceuticals; Automobile Manufacturers; Diversifying Banks
2	Movie & Entertainment; Airlines; Internet & Direct Marketing Retail; Pharmaceuticals; Hotels, Resorts & Cruise Lines	Movie & Entertainment; Oil & Gas Exploration & Production; Restaurants
3	Automobile Manufacturers	Apparel, Accessories & Luxury Goods; Hotels, Resorts & Cruise Lines
4	Oil & Gas Exploration & Production	Internet & Direct Marketing Retail

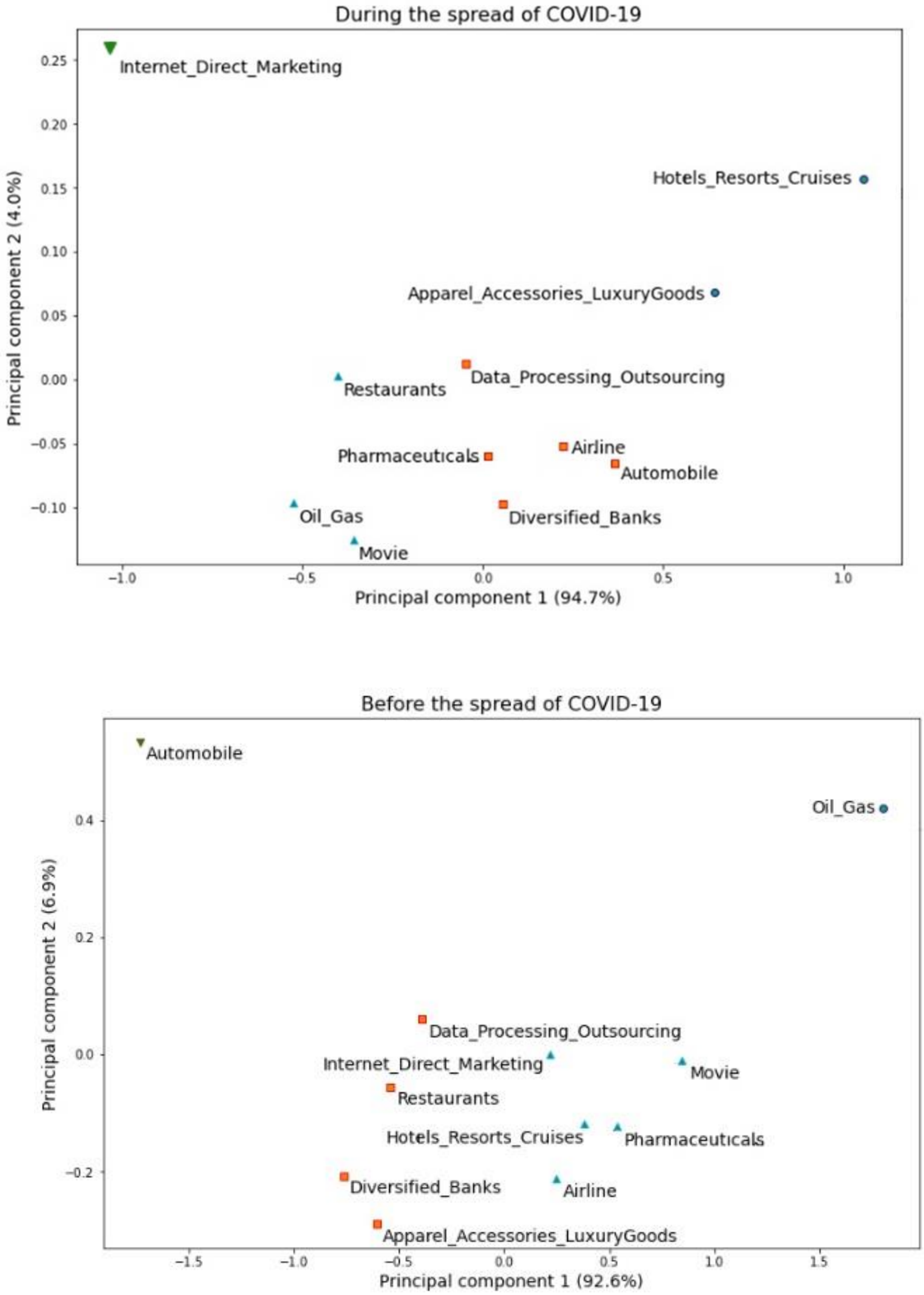

FIGURE 8 - Score plots of the two principal components of the conditional volatility hierarchical cluster solution. Maps for the daily S&P 500 industry indices before (top) and during (bottom) the spread of COVID-19

The Government and Central Banks play a key role in lessen the COVID-19 impact on stock prices, through fiscal and macro financial policies. In a period characterized by uncertainty, a reduction in interest rates would help to boost the investment and economic growth. Furthermore, a defer in loan repayments of firms in sectors most affected by COVID-19, would be a good way to reduce the losses of those companies and would consequently minimize the impact on the stock prices within those sub-industries. Furthermore, a boost in vaccine investment, increasing its availability and accessibility, would be a good way to decrease the uncertainty and make the stock market thrive. Additionally, a temporary decrease or even exemption from custom duties to companies in sectors affected by COVID-19 very dependent of imports, would be a good financial policy to minimize the company costs.

In order to measure the impact of the COVID-19, we next compute S&P500 index forecasts with fitted TGARCH models over the selection of cuttoff points and forecast horizons described in Table II. The objective is to compare the performance of the stock index during the pandemic to how it would have been during a period with more stability. The mean absolute percentage error (MAPE) forecasts evaluted for 1 to 10 steps ahead and for 23 cutoff points are presented in Figure 10 and 6, respectively. As expected, these MAPEs increase with the forecast horizon, however they look very different across industries. From Figure 10 it is possible to see that the pharmaceuticals industry is the one where the TGARCH forecasts revealed to be more accurate during the pandemic period, considering that the highest forecast error obtained is less than 3%. On the contrary, the industry that seemed to have been most impacted by the pandemic, considering the accuracy of the forecasts, is the Hotel industry, which has the highest average of mean absolute percentage error for the 10 step ahead forecast amongst all the sub-industries. The TGARCH model also revealed to have forecasted the Oil and Gas and the Airline industries less accurately when comparing the remaining sub-industries, which could mean that the COVID-19 pandemic had a bigger impact on these industries than on the other ones.

To understand which were the most problematic cutoff points in terms of the impact of COVID-19, we computed the third quartile (middle value between the median and the maximum) of each range of MAPEs per industry. For the 11 industries, In Figure 11, we can see that MAPE values for cutoff points 3, 4, 5, 6, 9, 10 and 21 are equal to, or greater than the third quartile threshold. Cutoffs 3, 4, 5 and 6 englobe the period between 19th February and 15th April. Among the reasons that could justify the higher impact of COVID-19 on stock prices during this period is the news related with the possible declaration of COVID-19 as a pandemic, which finally came to pass on the 11 March. Furthermore, on 13 March, the then U.S. President, Donald Trump, declared COVID-19 to be a National Emergency and banned non-U.S. citizens from travelling from Europe. Accordingly, some of this shocking news which came during a period with a lot of uncertainty associated with this disease could explain why these four cutoffs presented MAPE values that were higher than the usual ones. Hotels, Resorts & Cruise Lines and Oil & Gas Exploration & Production were the most destabilised sub-industries during the first months of the Pandemic. With regards cutoffs 9 and 10, which also revealed to be problematic in terms of the stability of stock prices, one possible cause could be the news of some emerging vaccines, such as that of AstraZeneca. Finally, in the last cutoff, which corresponds to the period between 3rd of November and 16th of November, the U.S. reported an unprecedented 100,000 cases in just one day, and Pfizer and Moderna revealed the results of their vaccine efficacy, which in effect means that this was a transition period in terms of confidence regarding the pandemic, which undoubtedly affected the stock market. The Hotels and Oil industries were the less stable sectors during this period.

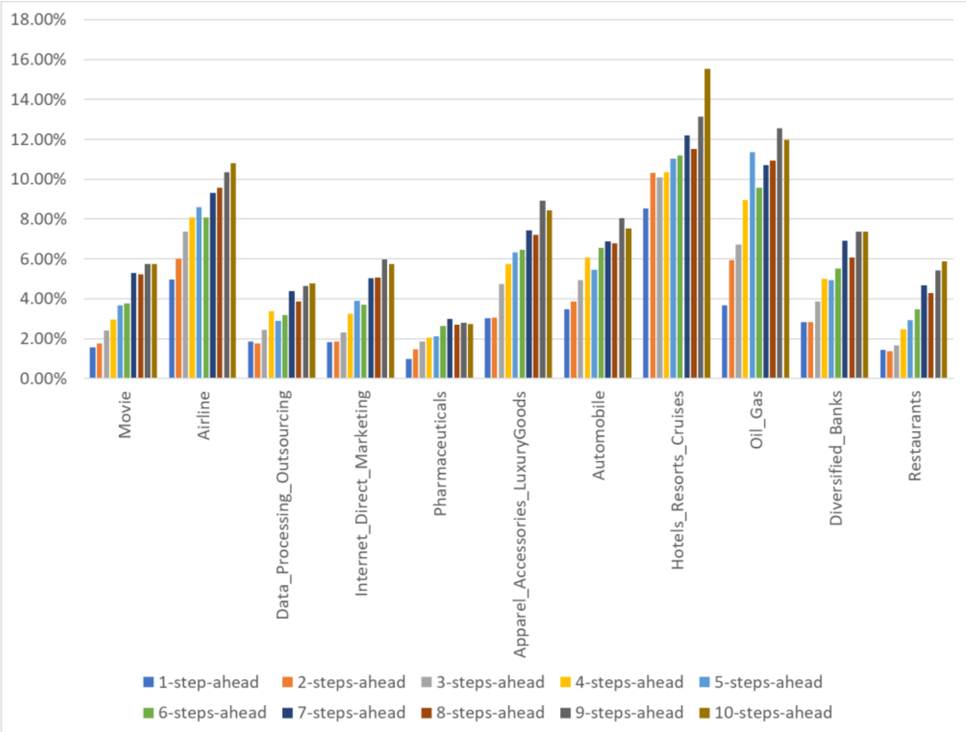

FIGURE 9 - 1 to 10 step-ahead forecast errors of TGARCH models for predicting S&P500 industry indices during COVID-19 pandemic

	Airline	Apparel	Automobile	Data	Diversified	Hotels	Internet	Movie	Oil	Restaurants	Pharm.
Cutoff 1	5,62%	7,63%	3,61%	1,20%	3,55%	7,70%	1,94%	1,73%	7,25%	1,82%	2,18%
Cutoff 2	4,32%	2,02%	2,64%	1,73%	3,04%	1,57%	4,33%	1,77%	4,90%	0,91%	1,20%
Cutoff 3	16,58%	9,27%	7,91%	9,18%	8,95%	16,82%	9,13%	7,42%	21,64%	6,71%	4,44%
Cutoff 4	16,20%	21,70%	20,65%	9,92%	21,27%	32,86%	7,18%	10,71%	56,83%	14,18%	4,56%
Cutoff 5	16,64%	7,76%	8,61%	5,55%	5,95%	9,76%	3,09%	6,23%	3,62%	8,31%	6,93%
Cutoff 6	14,08%	9,80%	8,44%	5,10%	5,63%	58,60%	5,58%	3,80%	14,90%	6,74%	4,57%
Cutoff 7	8,99%	4,80%	3,04%	2,84%	3,66%	4,14%	2,04%	1,21%	11,47%	4,44%	2,13%
Cutoff 8	22,42%	12,45%	7,33%	2,26%	10,89%	54,68%	2,28%	2,03%	3,78%	4,00%	1,38%
Cutoff 9	15,03%	9,17%	14,41%	5,31%	9,47%	9,03%	2,55%	5,79%	9,81%	5,00%	1,21%
Cutoff 10	14,53%	8,65%	8,08%	2,12%	6,53%	9,46%	3,79%	3,22%	11,50%	2,50%	1,03%
Cutoff 11	7,60%	2,99%	3,06%	2,65%	3,55%	4,18%	3,25%	3,70%	7,72%	1,54%	1,42%
Cutoff 12	3,06%	2,27%	1,45%	0,99%	5,80%	2,32%	6,21%	4,15%	2,31%	0,90%	1,70%
Cutoff 13	2,59%	4,21%	7,64%	1,30%	1,26%	2,56%	5,18%	4,94%	3,23%	2,49%	3,73%
Cutoff 14	2,60%	1,57%	1,67%	1,95%	1,40%	2,66%	2,42%	2,52%	4,33%	0,86%	1,16%
Cutoff 15	2,96%	3,26%	4,45%	1,07%	2,38%	3,96%	1,53%	1,28%	2,65%	1,95%	0,90%
Cutoff 16	9,31%	5,99%	3,98%	3,05%	3,72%	5,98%	2,86%	5,06%	6,57%	2,87%	0,91%
Cutoff 17	2,55%	5,28%	3,25%	1,98%	4,77%	4,08%	7,31%	3,46%	3,78%	1,04%	0,73%
Cutoff 18	2,10%	2,06%	2,44%	1,85%	1,89%	2,60%	4,23%	1,44%	1,73%	1,77%	0,44%
Cutoff 19	1,34%	3,04%	4,81%	1,17%	1,65%	3,07%	2,75%	2,35%	2,00%	0,49%	0,98%
Cutoff 20	4,19%	2,27%	4,94%	4,41%	1,97%	3,64%	2,41%	4,31%	3,02%	2,47%	2,38%
Cutoff 21	9,15%	5,13%	8,63%	7,27%	7,89%	13,12%	4,72%	6,28%	12,74%	4,15%	4,64%
Cutoff 22	3,76%	4,82%	3,48%	1,35%	2,95%	2,84%	0,98%	1,38%	7,92%	0,33%	0,90%
Cutoff 23	4,16%	4,09%	1,50%	1,29%	2,01%	4,25%	2,50%	2,29%	7,72%	0,92%	1,73%
3º Quantile	14,31%	8,20%	7,99%	4,75%	6,24%	9,61%	4,95%	5,00%	10,64%	4,29%	3,05%

FIGURE 10 - MAPE forecast errors for 23 cutoff points between 21st January and 11th December 2020 during the COVID-19 pandemic

From Table VII, it is possible to confirm that there is a negative moderate correlation between COVID-19 10-day-average new cases and the forecast errors. However, if instead we consider the growth factor in number of 10-day-average new cases, the correlation with the forecast errors ends up being positive and highly significant in most industries, including the Movie & Entertainment, Data Processing & Outsourcing Services, Pharmaceuticals, Apparel, Accessories & Luxury Goods, Automobile Manufacturers, Hotels and Resorts & Cruise Lines, Oil & Gas Exploration & Production, Diversified Banks, and Restaurants. It therefore appears that this growth factor in number of cases might be one of the causes that impact the accuracy of the forecasts. Since the correlations are positive for the growth factor, this means that the higher the growth in new Covid cases per day, the higher the MAPE values. This conclusion makes sense, considering the uncertainty generated by news related with a sharper growth in number of Covid cases, which, in turn, causes an increase in the volatility of stock prices, as previously stated by Baker et al. (2020).

Table VII: Pairwise correlations between COVID-19 (10-day-average new cases/growth factor) and TGARCH model forecast errors between 21st January and 11th December 2020)

Industry	Correlation between forecast errors and new COVID-19 cases	Correlation between forecast errors and COVID-19 growth factor
Movie	-0.228	0.701*
Airline	-0.424**	0.422**
Data_Processing_Outsourcing	-0.243	0.614*
Internet_Direct_Marketing	-0.322	0.288
Pharmaceuticals	-0.150	0.624*
Apparel_Accessories_LuxuryGoods	-0.374***	0.707*
Automobile	-0.303	0.690*
Hotels_Resorts_Cruises	-0.247	0.265
Oil_Gas	-0.233	0.730*
Diversified_Banks	-0.337	0.704*
Restaurants	-0.414**	0.837*

Note: *, ** and *** significant at 1%, 5% and 10% levels, respectively. While the correlation between forecast errors and new COVID-19 cases turned to be negative for all the sub-industries, the correlation between forecast errors and COVID-19 growth factor revealed to be positive and clearly significant for most of the sub-industries.

As mentioned before, a complete linkage (or furthest neighbor) is focused on the farthest pair of observations in two clusters. A single linkage (or nearest neighbor) concentrates on the shortest distance between a pair of observations in two clusters and a complete linkage (or furthest neighbor) is focused on the farthest pair of observations in two clusters. Both methods tend to group similar objects into clusters through a hierarchical agglomerative clustering procedure. From Figure 12, three clusters of industries were identified from the complete linkage dendrogram. One cluster is composed of Airline, Oil & Gas and Hotels industries, the second is composed by Automobile, Apparel and Accessories Luxury Goods and Diversified Banks, and the third cluster is composed of miscellaneous industries. The single linkage method identifies a clear outlier, namely the Hotels, Resorts & Cruises industry. This reinforces the fact that the Hotels industry was the one where the TGARCH forecasts of the stock indices were less accurate, considering it is classified as an outlier according to the single linkage method. The Airlines and Oil & Gas industries are grouped in a different cluster from the other sub-industries in both the complete and the single linkage dendrogram, because these two industries present much greater forecast errors when compared with the remaining industries, albeit in the 10 step the forecast error is not nearly as high as it is in the Hotels industry. The other industries that seem to have suffered a lower impact from the pandemic, considering their lower forecast errors are all grouped in the same cluster. The Automobile, Apparel and Accessories Luxury Goods and Diversified Banks industries all have similar forecast error values, which are all lower than the previously-mentioned industries, but higher than the remaining ones in the third group. According to the dendrograms, the sub-industries that experienced a lower COVID-19 impact in their stock indices, considering the TGARCH forecasts obtained, were the Restaurants, Pharmaceuticals, Data Processing & Outsourcing, and Movie and Internet Direct Marketing industries.

In order to better visualise and interpret similarities among industries, we constructed two-dimensional feature space maps of the complete linkage cluster solution. Figure 13 shows the corresponding 1-step-ahead versus 2-steps-ahead forecast errors and the 9-steps-ahead versus 10-steps-ahed forecast errors plots. It can be seen from these maps that the Hotels, Resorts & Cruises industry is clearly separated from any other industry in terms of short-term forecast accuracy, but it approaches the Oil & Gas and Airline industries in terms of the (in)ability to forecast in the long term.

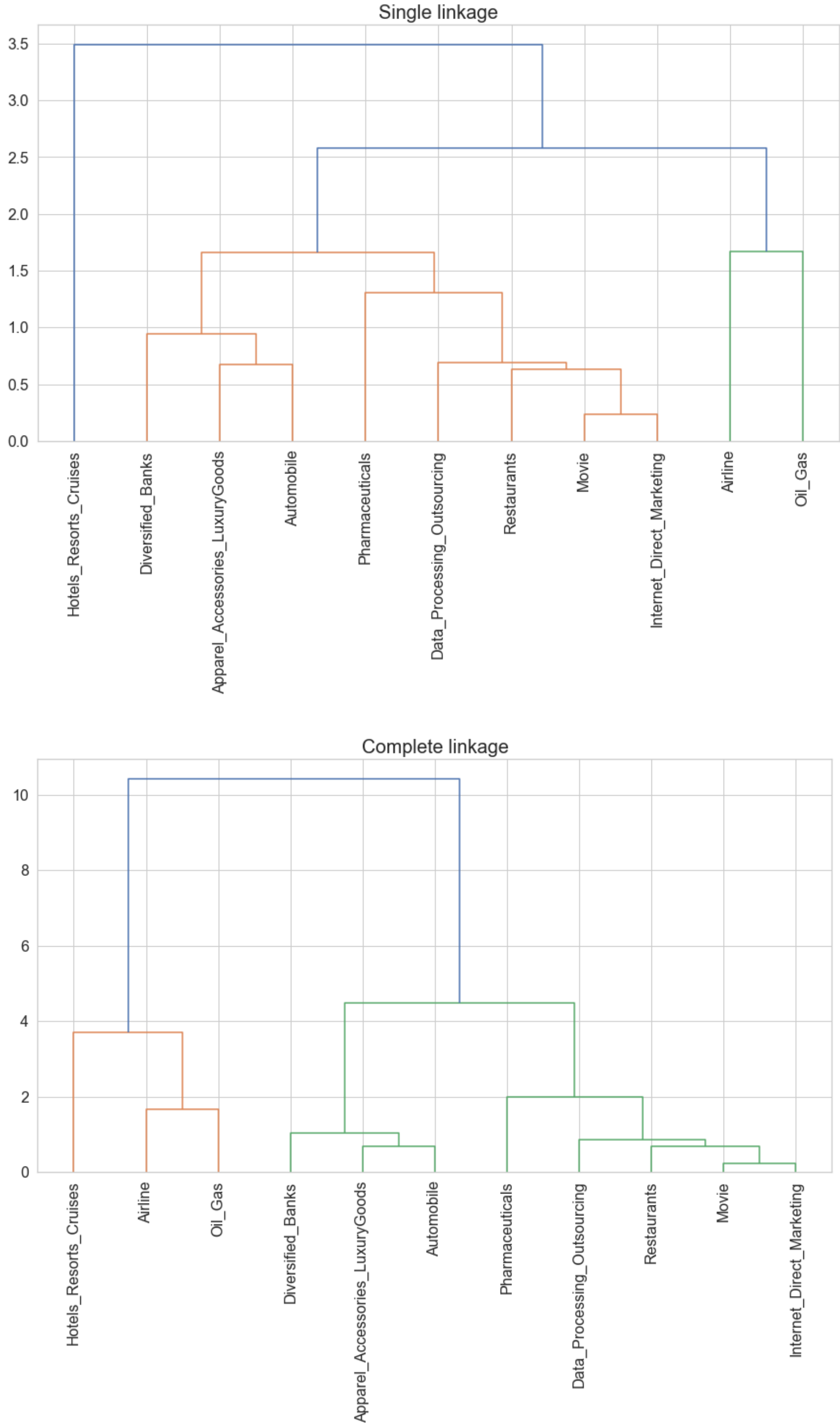

FIGURE 11: Single linkage (top) and complete linkage (bottom) dendrograms of S&P 500 industries, based on multi-step-ahead forecast errors during the COVID-19 pandemic

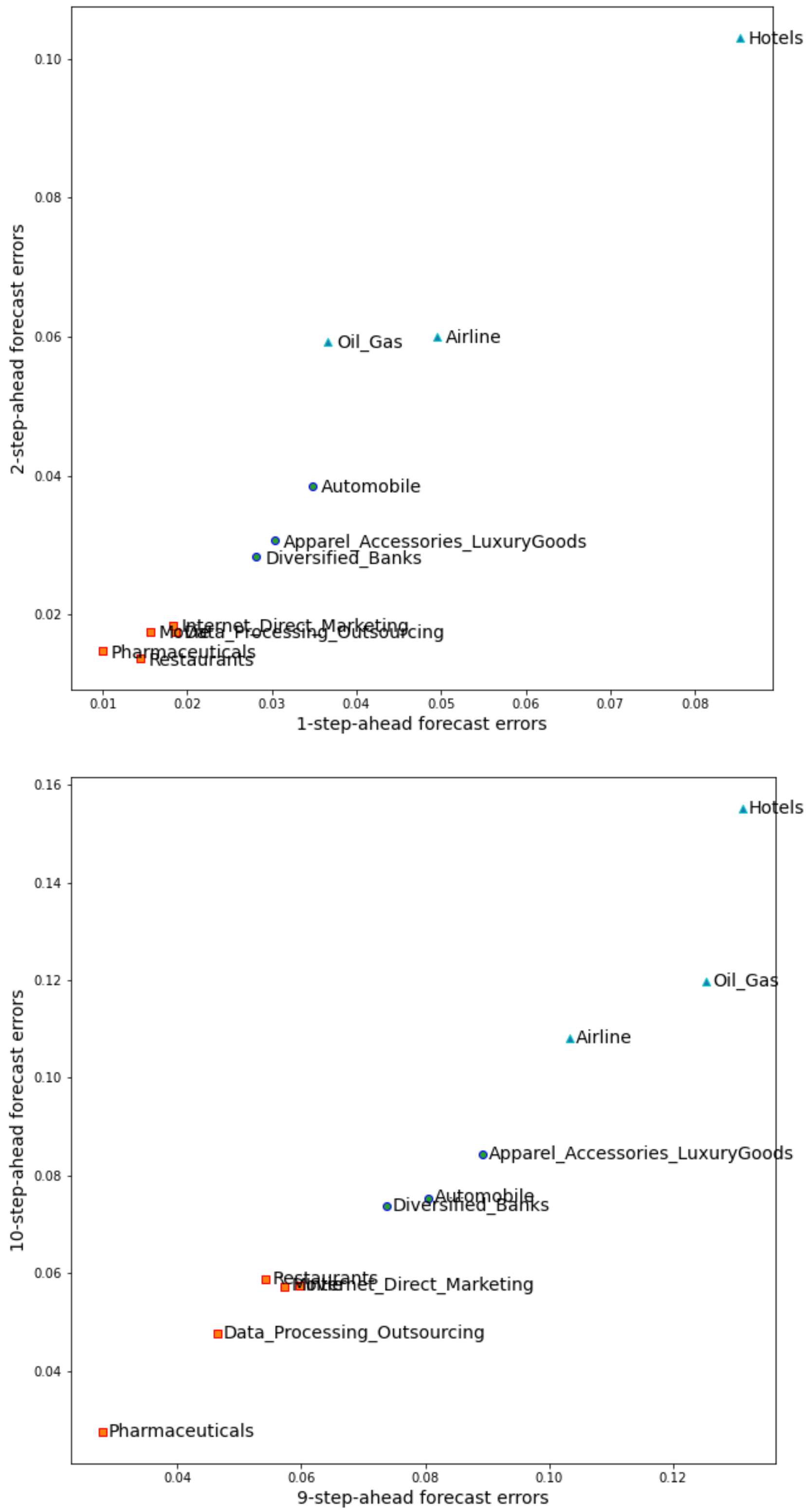

FIGURE 12**:** Two-dimensional feature maps of the S&P500 industry cluster solution: 1-step-ahead versus 2-step-ahead forecast errors (top), and 9-step-ahead versus 10 step-ahead forecast errors (bottom).

7. CONCLUSIONS

In this dissertation we investigated the impact of COVID-19 on U.S. stock market volatility with a focus on 11 S&P 500 industries, including Airlines, Automobile Manufacturers, Apparel, Accessories & Luxury Goods, Data Processing & Outsourced Services, Diversified Banks, Hotels, Resorts & Cruise Lines, Internet & Direct Marketing Retail, Oil & Gas Exploration & Production, Movies & Entertainment, Pharmaceuticals, and Restaurants.

For this purpose, we introduced a feature-based model method for clustering financial time series that accounts for the useful information regarding the dependence structure of their conditional volatilities. This clustering approach consists of fitting parsimonious threshold GARCH models to the industry stock returns and then computing the distance matrix between the autocorrelations of their estimated conditional variances. The method works for clustering short and unequally sized time series and can be applied in other time-varying variance time series, including exchange rates, interest rates, and commodities, among others.

By using hierarchical (complete linkage) and non-hierarchical (k-means) unsupervised clustering methods, we investigate the similarities and dissimilarities among the volatility of the S&P 500 industries before and during the spread of the COVID-19 pandemic. Both methods lead to identical and reliable results, however we found a very different cluster structure during the two periods (pre and post COVID-19).

Before the outbreak of Coronavirus, the Consumer Durables & Apparels, Financials, and Information Technology industries are grouped into one cluster and the Healthcare, Consumer Services, and Consumer Discretionary industries are grouped into another cluster. The Automobile and Oil & Gas industries are clearly separated from the rest in two isolated clusters. They also exhibit low/moderate pairwise correlations across volatility with other S&P 500 industries.

During the COVID-19 outbreak, except for the Internet & Direct Marketing Retail industry (which is booming), all industries experienced a significant increase in their volatility and approached each other in terms of time-varying variance. Among these, the industries most affected by the COVID-19 pandemic (Apparel, Accessories & Luxury

Goods, Hotels, Resorts & Cruise Lines) form a homogeneous cluster which is very close in Euclidean distance to two other industries impacted by the pandemic (Automobile and Airlines). All the industries became much more contemporaneous correlated in conditional variance with each other, with an emphasis on the Automobile and Oil & Gas, Movie & Entertainment and Banks, and Apparel, Accessories & Luxury Goods and Automobile industries.

Then, with the aim of using multi-step-ahead forecast errors of a well-known volatility model to identify similarities and dissimilarities among the S&P500 industries during the COVID-19 pandemic, we fitted simple threshold generalised autoregressive heteroskedasticity TGARCH(1,1) models to stock market indices to capture their volatility persistence and leverage (or asymmetric volatility) effects. Next, we calculated the mean absolute percentage errors (MAPEs) of 1 to 10 step-ahead forecasts for 23 different cutoff points from the beginning of the COVID-19 pandemic up until the end of 2020. Finally, we computed the standardised Euclidean distance matrix between each pair of industries and used the dendrogram and the two-dimensional feature maps associated with the underlying forecast errors to create distinct groups.

The industries most affected by the pandemic (Hotels and Airline) experienced a significant decrease in their relative forecast accuracy over the cutoff dates of 19th February - 3rd March, 4th March – 17th March, 1st April – 15th April, and 30th April – 13th May, during the first quarter of the COVID-19 pandemic crisis, in the year 2020. This forecast accuracy, measured for forecast horizons of 10 trading days, improved substantially over the study cutoff periods during the 2nd and 3rd Quarters of the Pandemic, in all of the S&P 500 industries.

The correlations between the multi-step-ahead forecast errors and the COVID-19 10-day-average new cases were negative and weak to moderate for all industries. On the other hand, the correlations between the forecast errors and the growth factor of this average are positive and significant in most industries (namely, Movie & Entertainment, Data Processing & Outsourcing Services, Pharmaceuticals, Apparel, Accessories & Luxury Goods, Automobile Manufacturers, Hotels and Resorts & Cruise Lines, Oil & Gas Exploration & Production, Diversified Banks, and Restaurants).

We found homogeneous clusters of industries in terms of the impact of COVID-19 on US stock market volatility. Two of the industries most affected by the pandemic (Airline and Hotels, Resorts & Cruise Lines) are grouped in a distinct cluster, together with Oil & Gas. Another two industries were impacted by the Pandemic (Apparel and Accessories & Luxury Goods and Automobile) and both form another distinct cluster, in conjunction with Diversified Banks. Both these two clusters are positioned far from those industries less impacted by COVID-19 and are typified by a better forecasting performance of the stock market, such as Pharmaceuticals, Movie & Entertainment, and Data Processing & Outsourcing Services. All the clustering methods used in this study lead to the conclusion that the industry that suffered the highest impact from the Pandemic was the Hotels and Resorts & Cruise Lines industry, which revealed to be a clear outlier in the clustering process, in the sense that its TGARCH forecasts were much less accurate than the forecasts from other industries.

Cluster analysis of time-varying time series using their stochastic volatility structure plays an important role in many financial applications, including portfolio diversification and asset allocation, the analysis of cross-country stock market co-movements, and the identification of similarities and dissimilarities among stock market industrial sectors and companies. The proposed model-based clustering method is simple, effective and can be applied to almost any kind of time series data.

The limitation of clustering approaches based on parameter estimates, residuals, and forecast errors of fitted models is the need for the ad-hoc modelling of several time series. Further research could explore the combination of model-based with features-based methods to investigate the degree of affinity of several financial time series, such as stock prices, market indices, exchange rates, inflation rates, and commodities prices, among others.

REFERENCES

Achille, A., & Zipser, D. (2020). A perspective for the luxury-goods industry during—and after—coronavirus. *McKinsey & Company*.

AJMC (2021). A Timeline of COVID-19 Developments in 2020. https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020

Albulescu, C. T. (2021). "COVID-19 and the United States financial markets' volatility". *Finance Research Letters*, 38, 101699.

Ashraf, B. N. (2020). Stock markets' reaction to COVID-19: Cases or fatalities?. Research in International Business and Finance, 54, 101249.

Ashraf, B. N. (2017). Political institutions and bank risk-taking behavior. *Journal of Financial Stability*, *29*, 13-35.

Baek, S., Mohanty, S. K., & Glambosky, M. (2020). "COVID-19 and stock market volatility: An industry level analysis". *Finance Research Letters*, 37, 101748.

Baker, S., Bloom, N., Davis, S., Kost, K., Sammon, M. and Viratyosin, T. (2020). "The Unprecedented Stock Market Reaction to COVID-19", the Review of Asset Pricing Studies, 10, 742–758.

Bastos, J. A. and Caiado, J. (2014). "Clustering financial time series with variance ratio statistics", *Quantitative Finance*, 14, 2121-2133.

Bastos, J. A. and Caiado, J. (2021). "On the classification of financial data with domain agnostic features", *International Journal of Approximate Reasoning***,** 138, 1-11.

Bureau of Transportation Statistics (2021). "U.S. Airlines 2020 Net Profit Down $35 Billion from 2019". https://www.bts.gov/newsroom/us-airlines-2020-net-profit-down-35-billion-2019

Caiado, J. and Crato, N. (2007). "A GARCH-based method for clustering of financial time series: International stock markets evidence", in: Recent Advances in Stochastic Modeling and Data Analysis (Edited by C. H. Skiadas), World Scientific Publishing, 542-551.

Caiado, J. and Crato, N. (2010). "Identifying common dynamic features in stock returns", *Quantitative Finance*, 10, 797-807.

Caiado, J., Crato, N. and Poncela, P. (2020). "A fragmented-periodogram approach for clustering big data time series", *Advances in Data Analysis and Classification* 14, 117-146.

Caiado, J., Maharaj, E.A. and D'Urso, P. (2015). "Time Series Clustering", in *Handbook of Cluster Analysis*, C. Henning, M. Meila, F. Murtagh, R. Rocci (eds.), 241-263, CRC Press, Taylor & Francis Group.

Carter, D., Mazumder, S., Simkins, B., & Sisneros, E. (2022). The stock price reaction of the COVID-19 pandemic on the airline, hotel, and tourism industries. *Finance Research Letters*, *44*, 102047.

Cepoi, C. O. (2020). Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil. *Finance Research Letters*, *36*, 101658.

Corporate Finance Institute (2022). "The S&P Sectors". https://corporatefinanceinstitute.com/resources/knowledge/finance/the-sp-sectors/

Chaudhary, R., Bakhshi, P., & Gupta, H. (2020). Volatility in international stock markets: An empirical study during COVID-19. *Journal of Risk and Financial Management*, *13*(9), 208.

D. Urso, P., De Giovanni, L. and Massari, R. (2016). "GARCH-based robust clustering of time series", *Fuzzy Sets and Systems*, 318, 56-79.

Deloitte (2020a). "Understanding COVID-19's impact on the automotive sector". https://www2.deloitte.com/global/en/pages/about-deloitte/articles/covid-19/understanding-covid-19-s-impact-on-the-automotive-sector

Deloitte (2020b). "Understanding the sector impact of COVID-19: Banking & Financial Markets". https://www2.deloitte.com/content/dam/Deloitte/global/Documents/About-Deloitte/COVID-19/COVID-19-Impact-Banking-Cap-Market-Sector.pdf

Deloitte (2020c). "Understanding the sector impact of COVID-19: Oil & Gas". https://www2.deloitte.com/za/en/nigeria/pages/finance/articles/understanding-the-sector-impact-of-covid-19-oil-and-gas.html

Ding, W., Levine, R., Lin, C., & Xie, W. (2021). "Corporate immunity to the COVID-19 pandemic", *Journal of Financial Economics*, 141, 802-830.

Engelhardt, N., Krause, M., Neukirchen, D., & Posch, P. N. (2021). "Trust and stock market volatility during the COVID-19 crisis". *Finance Research Letters*, *38*, 101873.

Engle R. and Patton, A. (2001). "What good is a volatility model?", *Quantitative Finance*, 1, 237-245.

Ertugrul, H. M., Güngör, B. O., & Soytas, U. (2020). The Effect of Covid-19 Outbreak on Turkish Diesel Consumption Volatility Dynamics.

Everitt, B., Landau, S. and Leese, M. (2001). Cluster Analysis, 4th ed., Edward Arnold: London.

Glosten, L.R., Jagannathan, R., Runkle, D.E. (1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", Journal of Finance 48, 1779-1801.

Granger, C.W.J., Ding, Z. (1996). "Varieties of long-memory models". Journal of Econometrics 73, 61-77.

Halkidi, M., Batistakis, Y. and Vazirgiannis, M. (2001) "On Clustering Validation Techniques", *Journal of Intelligent Information Systems*, 17, 107–145.

Hutchins Coe, E., Enomoto, K., Finn, P., & Weber, K. (2020). "Understanding the hidden costs of COVID-19's potential impact on US healthcare", Mckinsey & Company.

Investopedia (2021). "Global Industry Classification Standard (GICS)". https://www.investopedia.com/terms/g/gics.asp

Krishnan, V., Mann, R., Seitzman, N., & Wittkamp, N. (2020). "Hospitality and COVID-19: How long until 'no vacancy'for US hotels?", Mckinsey & Company.

Maharaj, E. A., D'Urso, P. and Caiado, J., (2019). *Time Series Classification and Clustering*, CRC Press, Taylor & Francis Group, United States.

MarketWatch (2021). *Market Data Center* [Database], New York, https://www.marketwatch.com/market-data

Mazur, M., Dang, M., & Vega, M. (2021). "COVID-19 and the march 2020 stock market crash: Evidence from S&P1500". *Finance Research Letters*, *38*, 101690.

Mittal, S. and Sharma, D. (2021). "The Impact of COVID-19 on Stock Returns of the Indian Healthcare and Pharmaceutical Sector", Australasian Accounting, Business and Finance Journal, 15, 5-21.

Onali, E. (2020). "Covid-19 and stock market volatility". Available at SSRN: https://ssrn.com/abstract=3571453 or http://dx.doi.org/10.2139/ssrn.3571453

Otranto, E. (2008). "Clustering heteroskedastic time series by model-based procedures", Computational Statistics & Data Analysis, 52, 4685-4698.

Otranto, E. (2010). "Identifying financial time series with similar dynamic conditional correlation", Computational Statistics & Data Analysis, 54, 1-15.

Papadamou, S., Fassas, A., Kenourgios, D., & Dimitriou, D. (2020). "Direct and indirect effects of COVID-19 pandemic on implied stock market volatility: Evidence from panel data analysis", MPRA Working Paper 100020, University Library of Munich.

S&P Dow Jones Indices (2021), *Global Industry Classification Standard (GICS) [Online]*. Available from: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-gics.pdf .

Simple Flying (2019). *Airline Capacity vs Airline Frequency – What is Better?* [Online]. Available from: https://simpleflying.com/airline-frequency-vs-capacity/

Song, H. J., Yeon, J. and Lee, S. (2021). "Impact of the COVID-19 pandemic: Evidence from the US restaurant industry". *International Journal of Hospitality Management*, 92, 102702.

Statista (2021a), "Weekly flights change of global airlines due to COVID-19 as of January 2021" https://www.statista.com/statistics/1104036/novel-coronavirus-weekly-flights-change-airlines-region/

Statista (2021b). "Disney+ subscriber numbers worldwide 2020-2021" https://www.statista.com/statistics/1095372/disney-plus-number-of-subscribers-us/

Statista (2022a), "Development of e-commerce shares in total retail sales in selected countries before and after the coronavirus (COVID-19) pandemic as of January 2021" https://www-statista-com.libdata2015.hilbert.edu/statistics/1228660/e-commerce-shares-development-during-pandemic/

Statista (2022b). “Netflix subscribers count worldwide 2011-2022” https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldrawide/

Wang, J., Lu, X., He, F., & Ma, F. (2020). Which popular predictor is more useful to forecast international stock markets during the coronavirus pandemic: VIX vs EPU?. *International Review of Financial Analysis*, *72*, 101596.

Zakoian, J.-M. (1994). “Threshold heteroskedastic models”, Journal of Economic Dynamics and Control 18, 931-955.