A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

Industry Preference of Retail Investors during Covid-19: Evidence from Robinhood

Francisco Rosa Faria
Mendonça Freire

Work project carried out under the supervision of:

Professor Daniele D'arienzo

16-12-2022

Abstract

This paper analyses how retail investors of Robinhood rearranged their portfolios across industries because of Covid-19, and what factors influenced their investment decisions. The analysis includes stock liquidity, historical beta, information display on Robinhood, and large one-day price movements to understand the changes in industry allocation. Retail investors, compared to a value-weighted portfolio, underinvested in industries with increasing market value following the Covid-19 Dow Jones crash, and overinvested in those with decreasing market value. A preference for high liquidity and low Price/Earnings-ratio stocks explains above 20% of the change in industry allocation, except in the *Energy* and *Consumer Discretionary* industries.

Keywords: Retail investors, Covid-19, industry, attention-induced trading, Robinhood

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

The Covid-19 outbreak was an extremely rare event which resulted in unprecedent consequences to social life, the economy, and financial markets. While disease outbreaks are not a new phenomenon, the consequences of this one are. The first stay at home order in the U.S. was issued in California on the 19th of March 2020, which closed non-essential businesses and restricted social gatherings. Soon after, many other states followed suit. These lockdowns, which imposed large restrictions on social and economic activity, had never been seen to this extent and scale prior to 2020. Likewise, the consequences for financial markets were unique for a pandemic. The Covid-19 pandemic stands out from the others for an extremely high frequency of large daily stock market moves that can directly be attributed to the pandemic (Baker et al. 2020).

Due to the nature of the lockdowns, the financial impact of the pandemic was not uniform across industries. Some industries, like air travel and hospitality, saw their economic activities extremely restricted, while others, like the consumer staples sector, were much less restricted. Baek et al. (2020) showed that the largest increase in total risk (measured as stock market volatility) was in the petroleum and natural gas industry, restaurants, hotels and lodgings industry, while there was a smaller impact on the food production industry, and beer and liquor industry. Financial investors reacted in accordance with this scenario, rearranging their portfolios across industries. Mazur et al. (2021) found that natural gas, food, healthcare, and software stocks earned the highest positive returns, while equity values in petroleum, real estate, entertainment, and hospitality sectors experienced a sharp fall, implying that investors fled from the latter industries, seeking refuge in the former industries. This reallocation was, however, not uniform across the types of investors. If some investors are selling in droves particular types of stocks, there must investors on the other side of the trade buying those same stocks. Glossner et al.

(2020) has shed some light on this question, analyzing how different types of investors traded around the Covid-19 pandemic. It found that institutional investors rearranged their portfolios towards financially strong companies, while hedge funds sold securities more indiscriminately. On the other hand, retail investors acted as liquidity providers, being more drawn towards those stocks that institutional investors were fleeing away from. It would, therefore, be expectable that retail investors rearranged their portfolios across industries differently from the market trend.

In this paper we study how retail investors from the Robinhood (RH) brokerage company reallocated their portfolio across industries amidst the Covid-19 pandemic, and what factors explain this behavior. Previous research has shown that retail investors behave differently from and are moved by different factors than institutional investors. Barber et al. (2006) has classified them as "noise traders"- traders who decide on buy and sell trades without the support of advanced fundamental or technical analysis- engaged in herding behavior. Recently, the retail investor market has changed considerably, with the appearance of brokerage companies such as RH, which doesn't charge brokerage fees, and offers a mobile-first app with features that make investing more similar to gambling (Popper, 2020). Due to its features, RH investors are particularly different from the rest of investors and retail investors, with half of RH users being first-time investors ("Robinhood raises $280 million", 2020), and have shown behaviors in many ways opposite to that of other investors. Barber et al. (2020) showed that RH investors are mostly driven by attention-based buying and herding behavior, being drawn to extreme gainers and losers, and susceptible to the display of information on the RH app. We, therefore, want to analyze whether RH users invested in different industries than the market trend, and whether attention-based buying, large return gains and losses, and the display of the information on the app was behind their portfolio reallocation between industries, amidst the Covid-19 pandemic.

We will initiate this paper by providing a general literature overview of the present conclusions that have been obtained from datasets on retail investors and from the Robintrack (RT) dataset. From this, we will identify the most relevant behaviors of these investors that will be used to attempt to explain the variation in industry allocation. We will further present the most relevant research on investor behavior during Covid-19 and other financial uncertainty shocks. Afterwards, we will present a graphical analysis to compare the evolution of aggregate holdings for each industry with their respective aggregate market value, which will show if Robinhood users moved in accordance with the market trend, or were contrarian forces, in each industry. Finally, using a pooled OLS and Fixed Effects linear regression, we will analyze which behaviors and factors are most relevant in explaining the variation in industry allocation during the Covid-19 pandemic.

2. Literature Review

Through the 1990's, research of investor performance and behavior was almost exclusively focused on institutional investors, mostly due to data availability. Early research on retail investors have classified them as noise traders with sub-par performance. Barber et al. (2006) showed that individual investors trading at a discount brokerage and at a retail brokerage were systematically correlated, meaning that, in any given month, they tend to buy and sell the same stocks. Odean (1999) analyzed the trading records of investors from a discount broker between 1987 and 1993 and concluded that these investors systematically earn subpar returns before costs.

More recently, there's been research on retail investors based on the RT dataset. These papers add to literature by including the Covid-19 market crash, since "earlier studies did not include precipitous market declines" (Welch, 2020). In "The Wisdom of the Robinhood Crowd," Welch (2020) concluded that RH investors bought stocks when these experienced large increases or decreases in price. They did not panic in the February/March 2020 Covid market drop, having increased their relative holdings in individual stocks, as well as their overall holdings. These is in accordance with earlier research (Grinblatt and Keloharju (2001) and Barber and Odean (2008)), which had shown that retail investors in the 1990s had bought stocks with recent large price up or down swings. However, contrary to the aforementioned research, it showed that the RH portfolio performed well for the analyzed period, earning positive alphas in various models. It also has shown that they have some interest biases (i.e., overweighting stocks of companies and products their demographic has more interest in). Of importance to our paper, in the choice of variables for analysis, it showed that RH investors were most attracted to stocks with high trading volume in the last 12 months. Glossner et al. (2020) also studied RH users' investment decisions during Covid-19, comparing it to that of institutional investors. It showed that institutional investors increased price crashes, net selling stocks and seeking refuge in companies with strong financials (high-cash and low-debt firms). Similar to Welch (2020), it concluded that Robinhood users had a contrary reaction, as retail interest surged for firms with high leverage and low cash holdings, acting as market stabilizing force. More research has been done on the investment behavior of RH traders. Moss et al. (2020) studied the importance of ESG announcements for their choice of stocks, showing that portfolio changes related to ESG press releases are not statistically different than the adjustments that occur on non-event days; rather, the changes are economically significant in response to other press releases, especially those pertaining to earnings announcements. Barber et al. (2020) showed that RH investors are mostly driven by attention-based buying and herding behavior, being drawn to extreme gainers and losers, and susceptible to the display of information on the Robinhood app, as they were more likely to buy stocks in the app's "Top Movers" list, showing the stocks with largest absolute changes in price. Previous research on this behavior by Barber and Terrance (2005) had shown that that attention-based buying could be a result of the problem of too many choices in the investment universe, which implies high search costs.

Other research related to this paper includes the response of investors and financial markets to the Covid-19 crisis, as well as other shocks and periods of great uncertainty. A large set of this research focuses on investor behavior and psychology, namely how investor sentiment and changes in risk aversion led their choices and affected the market. R. Ortmann et al. (2020) focused on retail investors and the changes to their trading patterns and risk taking during the Covid outbreak. By doing an OLS regression on variables such as number of trades in a week, leverage employed in a trade (as a proxy for risk taking), and the number of covid cases per day, they concluded that the intensity of trading increased with the pandemic (compared to previous) and even more after the Dow drop, in accordance with Welch (2020). Furthermore, it showed that investors were more risk averse, using less leverage, but nonetheless, did not move to "safe-haven" stocks, such as gold and government bonds. These findings contrast with investors' reactions to other shocks marked by great uncertainty, such as terrorist attacks, which are associated with reduced flows to risky asset classes (Wang and Young, 2020) and heavy retail investor selling (Burch et al. 2016).

Of particular interest to this paper, some research, regarding Covid-19, has been done on the industry-level. Baek et al. (2020) studied the impact of the pandemic on US stock market volatility. It concluded that, while increases in total and idiosyncratic risk were observed across all industries, changes in systematic risk varied across industry. The largest increase in total risk is in the petroleum and natural gas industry, and restaurants, hotels and lodgings industry, while there was a smaller impact on the food production industry and beer and liquor industry. Mazur et al. (2021) studied the performance of the US stock market during the Covid-19 market crash, including an industry level analysis. It found that natural gas, food, healthcare, and software stocks earned the highest positive returns, while equity values in petroleum, real estate, entertainment, and hospitality sectors experienced a sharp fall.

In general, the research on investor behavior during a market downturn focuses on a macro level, by studying how stock exchanges indexes move, or focuses more on the psychology of their choices, by analyzing their risk aversion and intensity of trade. The research on retail investors and those using the RT dataset analyze their performance as well as what is associated with the buying and selling of given stocks. This paper adds to the literature by providing an industry level analysis of how the Covid outbreak changed the holdings portfolio of RH users and what is behind or associated with the users' trading decisions. We build on previous research which identified important variables and the relevant investor psychology that explain the RH holdings, to test their role in the reallocation of holdings across industries amidst the Covid-19 outbreak.

3. Data

3.1 Robintrack

The main dataset used was sourced from the Robintrack.net website. Starting in 2018, RT downloaded publicly available RH data, collecting the number of user accounts which held each stock at each observation. The data goes until August 2020, when it was shut down by RH, which was concerned that third parties could use the data against their own clients (Ponczek 2020). The RT has information on 8597 securities over 802 day-observations, beginning on May 5, 2018 and ending on August 13, 2020. For each, we extracted the last observation of each day.

Some tickers don't appear at the start of the RT data, while others do not appear at all, as is the case of CELG (Celgene) and TMX (Time-Warner) (Welch 2020). As can be seen in Figure 1, the level of missing values decreases with time. Furthermore, some tickers had observations for all days, however, they started as “0”, than jumping to a very high number, comparatively; others started with normal values, but then showing zeros either until the end, or in the middle of normal values (that is, breaking the trend for a few observations, signaling that it is a missing value). Of the 8597 securities, only 5288 had observations for all the days in the data. Since my analysis concerns the events around Covid, the analysis will be restricted to after the first day of 2019, and the regression analysis will be restricted to after the 14th of August 2019 to the 13th of August 2020 (inclusive). This serves to both restrict the missing values and a time bias for the estimation of the pre-Covid coefficients. For this sample, there are 6144 tickers without missing values.

The RT data also have missing days. On March 2, 2020, and March 9, 2020, the website was subjected to system wide outages; on August 9, 2018, on January 24–29, 2019 and January 7–15, 2020 the script failed to run (Welch 2020). Overall, there were 12 business days missing. On the other hand, many day-observations have repeated information (i.e., no change regarding the previous observation). This is due to weekends and holidays when the stock markets close. Eliminating those, there are 397 observation-days.

3.2 Yahoo Finance

The quote type of each stock was obtained by taking advantage of a Python API (pandas_datareader) which extracts information from the Yahoo Finance website (YF). Furthermore, to match the tickers to the Eikon platform, the stock exchange on which each security was quoted was needed; this information was also obtained from YF. Of the total RT securities, 7541 were matchable to YF. Of the 6144 securities with no missing values, we are left with 5619. The missing securities are mostly due to companies that changed their ticker since the data were compiled.

3.3 Eikon

The rest of the data (market value, trading volume, historical beta, historical price and industry code) were obtained from the Refinitiv Eikon platform. Of the 7541 securities found on YF, 668 were not matchable. This leaves us with 5057 securities without missing RT values. The investment universe we will be analyzing will concern only these securities. The data on industry also have many tickers without industry, which is due to the presence of ETFs. Even though they are of no relevance to our study of industry, they are relevant to accurately estimate the control variables, and therefore will not be eliminated from the sample.

3.4 Final Dataset

In this section, we will discuss the final dataset, after all data preparation, as described above, was performed.

Our investment universe encompasses 5057 securities, from January 2, 2019, to August 13, 2020, which constitutes 2,007,629 day-observations. Most securities, 3559, are equity stocks, with 1498 being ETFs (Figure 2). There is also some imbalance in the distribution of industry (Figure 3). *Financial* stocks comprise most of the investment universe, being 6 times more than each of the three smallest industries. *Utilities* and *Telecommunications* are the two smallest, with 108 and 109 securities, respectively. Descriptive statistics can be seen in Table 1.

4. Hypothesis and Methodology

4.1 Hypothesis outline

The main objective of this paper is to understand the impact of Covid-19 on the preferences retail investors from a large discount broker show for stocks of different industries. Previous research has been able to identify behaviors of retail investors in general, as well as of RH users, and it has shown that they have some interest biases (i.e., overweighting stocks of companies and products their demographic has more interest in). In this paper, we will analyze whether RH users show preferences in their choice of industry, how this has changed following the Covid-19 outbreak, and whether these changes persist after controlling for the influence of attention induced trading, the display of information on the Robinhood app, and investors' risk-aversion levels. For this analysis, we will have to measure the allocation of holdings across industries before and after the outbreak, which requires the establishment of an appropriate event day to distinguish between different periods. Since the subject of the study concerns financial markets, the most theoretically sound day would be February 20, 2020, when the uncertainty around Covid culminated in the Dow Jones experiencing its largest intra-day crash since the first case was confirmed.

4.2 Empirical Model and Variable Definition

To test the hypothesis that RH users were attracted differently to stocks of each industry and that their preference of industry was significantly altered during the Covid-19 outbreak, two methodologies will be used. First, a graphical analysis will be performed, to analyze how the allocation of holders per industry changed across time, and how it relates to aggregate market value of the sample stocks belonging to each. After, a linear regression will be performed, to control for other variables that have been shown to influence RH users, and thus estimate how much of the inter-industry holding variation can be explained by already established behavior of RH users and retail investors. For the first regression, the following model is specified:

$$\text{logHolders}_{i,t} = \alpha_i + \beta_1\text{Industry}_i + \beta_2\text{Industry}_i\times\text{Covid}_t + \beta_3\text{logTradingVolume}_{i,t-1} + \beta_4\text{logTradingVolume}_{i,t-1}\times\text{Covid}_t + \beta_5\text{logPriceEarningsRatio}_{i,t-1} + \beta_6\text{logPriceEarningsRatio}_{i,t-1}\times \text{Covid}_t + \beta_7\text{TopList}_{i,t-1} + \beta_8\text{TopAbsoluteReturn}_{i,t-1} + \beta_9\text{Beta}_i + \beta_{10}\text{Beta}_i\times\text{Covid}_t + \varepsilon_{i,t} \quad (1)$$

The model uses one dependent variable:

- $\text{logHolders}_{i,t}$, representing the holders of stock *i* at time *t* as a percentage of aggregate number of holdings of all stocks, log transformed.

For the dependent variables, we have used variables that previous studies using the RT dataset have already shown to be relevant to Robinhood investors and explain the allocation of investment. Based on Welch (2020), the number of shares traded over the last year was added as $\text{TradingVolume}_{i,t}$. According to that same paper, the variable $\text{TopAbsoluteReturn}_{i,t}$ was included, indicating the number of times a stock had large price swings over a period of 4 weeks. The variable $\text{TopList}_{i,t}$ is a proxy the number of times stock *i* appeared on the app's Top Movers list over a period of 4 weeks. Since there is no historical data for that list, it was created by including the stocks with the most changes in price in the dataset and with a market value larger than 300 million at the respective date (a requirement to appear on that list); this will not be the exact list that would have appeared on the app, however, it is expected to be a close approximation. This variable was added based on Barber et al. (2020), which showed that RH investors are susceptible to the display of information on the app and are more likely to invest in stocks on the Top Movers list than stocks with similar price movements not eligible for the list. Price earnings ratio was included due to being information that RH provides in the details page of a stock ("Viewing stock detail pages.", n.d.). Finally, the historical beta of each stock was included to control for investors' risk-aversion levels. The beta and trading volume were interacted with a Covid dummy since it is reasonable to assume that investors risk-aversion levels and preference for liquid securities changed with the financial turmoil. Hence, the explanatory variables are:

- Industry_i, the company's general industry classification based on the Industry Classification Benchmark launched by Dow Jones and FTSE (*Financials*, *Health Care*, *Consumer Discretionary*, *Industrials*, *Technology*, *Energy*, *Basic Materials*, *Real Estate*, *Consumer Staples*, *Telecommunications*, *Utilities*), which will be one hot encoded into 11 dummy variables;

- $Industry_i \times Covid_t$, the dummy variable of the company's general industry classification, which takes value of 1 starting February 20, 2020, and 0 before;
- $logTradingVolume_{i,t-1}$, the number of shares traded of stock *i* over the previous 52 weeks, at time *t-1*, divided by the average number of shares traded of all sample stocks over the previous 52 weeks, at time t-1, log-transformed;
- $logTradingVolume_{i,t-1} \times Covid_i$, being equal to $logTradingVolume_{i,t-1}$ for observations after the 19th of February, 2020 (not inclusive), and zero otherwise;
- $logPriceEarningsRatio_{i,t-1}$, price divided by the earnings rate per share of stock *i*, averaged over the previous 6 months, divided by the average of all sample stocks at time *t-1*, log transformed;
- $logPriceEarningsRatio_{i,t-1} \times Covid_i$, being equal to $logPriceEarningsRatio_{i,t-1}$ for observations after the 19th of February, 2020 (not inclusive), and zero otherwise;
- $TopList_{i,t-1}$, the number of times the stock would have appeared on the app's Top Movers list according to the proxy for the list, in the 20 business days before time *t*;
- $TopAbsoluteReturn_{i,t-1}$, the number of times security *i* was in the top 10 percentile and bottom 90 percentile of day returns in the 20 (business) days prior to day *t*. Returns are calculated as $\left(\frac{P_t}{P_{t-1}} - 1\right)$, with P_t being the closing price of day *t*;
- $Beta_i$, the historical beta of stock *i*, calculated at the start of the regression sample, August 14, 2019;
- $Beta_i \times Covid_i$, being equal to $Beta_i$ for observations after the 19th of February, 2020 (not inclusive), and zero otherwise.

Prior to obtaining the results, it was necessary to check that the classical linear model assumptions hold to unbiasedly estimate p-values and t-statistics. First, there doesn't seem to be any strong indication of problems with multicollinearity. As can be seen in the Table 2, there is no correlation above 0.5 (in absolute value) between the dependent variables in the sample. To aid in the diagnosing of multicollinearity, each independent variable was regressed on all other independent variables, using the Variance Inflation Factor (VIF) as a metric. The VIF is obtained by the following formula: $VIF = \frac{1}{1-R2}$. The typical cut-off values are 5 for moderate and 10 for strong multicollinearity. The results (Table 3) show that there aren't any strong or moderate multicollinearity problems. Trading volume also includes retail investor participation and, therefore, there is a multicollinearity concern of this variable with the dependent variable. However, first, RH users represent a small subset of trading- during Robinhood outages, trading in stocks dropped only 0.7% (Barber et al. 2020)-, and therefore this problem would be small; second, according to Welch (2020), "the data further suggests that it is the retail investors who disproportionately end up owning these highly liquid stocks." A well-documented behavior of retail investors and of RH users is attention induced trading (Barber et al. 2020), and trading volume is a reasonable proxy measure of a high attention stock. Hence, there is strong theoretical evidence to not eliminate trading volume as a control variable.

There is also evidence of the violation of other assumptions (Figures 4-6), namely, serial correlation of errors and homoskedasticity. To correct for this, the Newey-West estimator with 21 lags (based on the autocorrelation plot in figure 6) will be used for calculating the t-statistics.

The model described above will first be run using Pooled OLS. This method, however, raises the concern of unobserved time constant effects not being accounted for and leading to correlation among error terms. Since the unobserved effects might also be correlated with the explanatory variables, it can lead to biased and inefficient estimators. To avoid this problem, the model will also be run using a Fixed Effects Estimation which uses time-demeaned variables (removing from each variable the average of each stock) to eliminate the unobserved time constant effect α_i (Wooldridge 2012). As a result, the variable Industry_i and Beta_i will also be eliminated, since they are constant across time:

$$\text{logHolders}_{i,t} = \beta_1\text{Industry}_i\times\text{Covid}_t + \beta_2\text{logTradingVolume}_{i,t\text{-}1} + \beta_3\text{logTradingVolume}_{i,t\text{-}1}\times\text{Covid}_t + \beta_4\text{logPriceEarningsRatio}_{i,t\text{-}1} + \beta_5\text{logPriceEarningsRatio}_{i,t\text{-}1}\times\text{Covid}_t + \beta_6\text{TopList}_{i,t\text{-}1} + \beta_7\text{TopAbsoluteReturn}_{i,t\text{-}1} + \beta_8\text{Beta}_i\times\text{Covid}_t + \varepsilon_{i,t} \quad (2)$$

To test the hypothesis, we will use changes in R^2, and absolute value of coefficients and t-statistics from variable omission (Grömping, 2015). The above models will be tested in different ways. First, we will only use the industry variables and run 2 regressions:

$$\text{logHolders}_{i,t} = \alpha_i + \beta_1\text{Industry}_i + \beta_2\text{Industry}_i\times\text{Covid}_t + \varepsilon_{i,t} \quad (3)$$

$$\text{logHolders}_{i,t} = \alpha_i + \beta_1\text{Industry}_i + \varepsilon_{i,t} \quad (4)$$

The difference in R^2 between the former and the latter will be the explanatory power of $\text{Industry}_i\times\text{Covid}_t$ in a model without controls.

Afterwards, we will run the regression with control variables in the same way: first with $\text{Industry}_i\times\text{Covid}_t$, and then without. The difference in R^2 between the former and the latter will be the explanatory power of $\text{Industry}_i\times\text{Covid}_t$ in a model without controls. By comparing the two R^2 differences, we will be able to establish how much of the explanatory power of $\text{Industry}_i\times\text{Covid}_t$ can be attributed to the control variables. Other measures of variable importance which will be used are the absolute value of the coefficients and of the t-statistics of $\text{Industry}_i\times\text{Covid}_t$. By comparing them between the model with and without control variables, we will be able to ascertain which industry changes in holdings are most explained by our model, and whether some persist. This way, we will be able to understand how much of the change in industry preference can be explained by our model.

Finally, we want to ascertain the importance of each control variable in explaining the change in industry aggregate holdings. For this purpose, we will run a regression with all control variables but for one, for each of the control factors. This will be done twice, first with $Industry_i \times Covid_t$ as a regressor and then without, and the same R^2 differences will be obtained. With this method, the higher the R^2 differences, the more important was that factor in accounting for the change in industry preference.

Since it is the change in industry preference because of Covid-19 that is under analysis, the regression sample will cover a smaller period than the one used for graphical analysis, from the 14th of August 2019 to the 13th of August 2020 (inclusive), so as to have the same number of observation-days prior to the Covid event day as after.

4.3 Limitations

There are several limitations to this analysis. First, we are restricted to the RT dataset, which poses limitations to external validity. RH, being a fintech investment app, serves a very particular demographic of retail investors. The average age of its users is 31, and about 50% of those are first-time investors (Lam, 2022). RH targets younger and less experienced investors, and, therefore, their behavior might not be representative of the average retail investor. Furthermore, most securities are American based, which leaves our analysis with a geographical bias.

Other limitations arise from problems with the dataset and RT. As described previously, many stocks only show up later in the data, while some days are missing or incomplete due to fails in the RT script. The tickers with incomplete data were eliminated from the sample, which poses the threat of biasing the results. For those missing values, it is expected that there would be no bias, since they are the result of technical problems in collecting the data and should, therefore, be missing completely at random. Nonetheless, they present a problem as our analysis does not present a complete picture of RH investing. On the other hand, there is no reason to believe the tickers that were not matchable to the Eikon platform are MCAR. However, including them through other methods other than eliminating them (e.g., imputing a median or eliminating them only for the day-observations missing) is not a necessarily better solution since the pre-Covid and post-Covid sample will be different, which might equally bias the estimations. Another limitation of the RT data is that it only shows the number of user accounts holding each stock, and it is not possible to know the value invested in each stock. This limits our investigation severely. For instance, we might conclude that users were more attracted to a stock after covid than before, because the number of accounts holding that stock increased, when in fact more investors are holding it but investing less money in it. Thus, the number of holdings can only be an imperfect proxy for the investment portfolio of RH users.

Finally, there is one concern that relates to a specific RH feature. Each new RH user, upon signing-up, receives one free share of a randomly chosen stock. While 2% of these new accounts receive a share from a pool of six stocks priced above $10, and these stocks are explicitly named on Robinhood's bonus page, it is impossible to identify the shares received by the other 98%. With these new holdings representing only one share, they can misrepresent their importance and bias the results.

5. Analysis of the RH Portfolio

5.1 Aggregate Holdings

Previous studies on retail investors had shown that they were likely to buy when they had large price changes. Grinblatt and Keloharju (2001) and Barber and Odean (2008) showed that retail investors in the 1990s had bought securities that had increased or decreased dramatically in price recently. These studies, however, have limited applicability for the case of sharp and prolonged market decline. In the Grinblatt and Keloharju (2001) sample there was no steep market decline, while the largest fall in the Barber and Odean sample was not above 15%. On the contrary, our sample includes the Covid bust and subsequent bull market. After the S&P 500 reaching its highest point of 3,386 on the 19th of February, 2020, it started declining, reaching its lowest point on March 4, 2020, a low 2,237. This represented a 33% fall, significantly larger than any decline present in the previously mentioned papers. Retail investors could behave differently in such a scenario; however, if they were to behave similarly, one would expect an increase in buying during the decline, as well as during the subsequent market boom. We will look at RH aggregate holdings of all stocks to analyze this behavior.

The aggregate number of holders of all stocks had been increasing from the start of the sample. Through 2019, it increased at a rate of 0.15% per (trading) day; this is contrasted with the sample for 2020, which experienced a growth rate of 0.74% per trading day. As can be seen in Figure 7, as the market suffered its large Covid decline, Robinhood investors increased their holdings at a faster rate, and kept increasing even in the following bull market. The Covid market drop coincided with an increase of RH holdings. Ozik, Sadka, and Chen (2020), analyzing the same RT dataset, which reached a similar conclusion, found a link between this acceleration and the Covid lockdowns, and suggested that retail investors were a major force in attenuating the rise in stock market illiquidity during the first months of the pandemic.

5.2 Industry Allocation

Robinhood investors did not invest equally across the different industries. Welch (2020) had already showed that there are several peculiarities regarding the RH users' choice of stocks, as they tend to overweight stocks that appeal more to the RH demographic. AMD, Snapchat and Cannabis stocks were especially popular stocks; for a moment at the end of 2019, the most held stock was Aurora Cannabis (ACB). This peculiarity can also be seen, to some level, on the industry allocation of RH users.

At the start of the sample, the most held industry, adjusted for the number of securities of that industry in the sample, was *Technology*, with 3.2 times the users that an equal weight portfolio (same number of holders for all stocks) would have (Figure 8). On the other hand, *Financials* was the least held industry, with around 41% the users of an equal weighted portfolio. Comparing with the value-weighted portfolio (Figure 9), the picture is similar, with slight differences. The most overweighted industry is *Consumer discretionary*, but *Technology* is a very close second. Those, along with *Healthcare* and *Basic Materials*, were overweighted by RH users.

Across the sample, and especially during the Covid outbreak, there were tremendous changes in the industry allocation. Most interestingly, these changes seem to be significantly different from the behavior of the market and other investors, indicated by how the market value of each industry moves across time. The industries that experienced growth between February 19 to August 13, 2020, were *Basic materials*, *Technology*, *Consumer Discretionary*, and *Healthcare* (Figure 10). With the exception of *Consumer Discretionary*, all these industries experienced a decline in the percentage of the sum of holdings during this period, and a steeper decline than previously. *Technology* even fell to being underweighted, compared to the value-weight portfolio, at the end of the sample. The behavior relating to *Consumer Discretionary* might, at first glance, appear to be opposed to that of those other industries. However, a closer look at the different subsectors is needed. *Consumer Discretionary* is divided into 5 subsectors, which are: *Consumer Products and Services*, *Travel and Leisure*, *Retail*, *Media*, and *Automobiles and Parts*.

All the subsectors are overweighted in the RH portfolio, compared to the equal weight portfolio (Figure 11) ; however, they do not behave similarly across the sample. Specifically, *Travel and Leisure* stands out as an extreme outlier, with its weight increasing almost seven-fold at the end of the sample, compared to the beginning. The second sector with largest increase was *Media*, which did not even see a two-fold increase. Curiously, *Travel and Leisure* was also the sector to experience largest fall in market value during the pandemic (Figure 13). The sectors that experienced market value growth following during the Covid section of the sample were *Retail*, *Consumer Products and Services*, and *Automobiles and Parts* (Figure 13). Like before, except for the latter, these sectors experienced a steeper decline in their weight relative to the equal weighted portfolio; even though *Automobiles and Parts* saw an increase, it became more underweighted compared to the value weighted portfolio. The opposite behavior can be seen in the industries that experienced a consistent fall in market value (as opposed to a large increase followed by a large decrease). With the exception of Real Estate, RH users overinvested in the industries that performed poorly.

Robinhood users' choice of industries to invest in was peculiarly different than the rest of investors and showed some interest biases. They had a slight tendency towards the industries that they had more interest in. Following the February 19 stock market crash, they showed less interest in the industries which strived and to which investors, on the aggregate, were attracted to, compared to the other industries. They also tended to move to the industries that suffered most.

6. Results

The results for the pooled OLS and Fixed Effects are described in Table 4. In column (1), pooled OLS was used, with no control variables. The coefficients for industry are all statistically significant, indicating that stocks of some industries have more holders than stocks of others, across the sample prior to the event day for the pandemic. *Technology* is the most preferred industry prior to the Covid event day, with a coefficient of 2.56, indicating that stocks belonging to that industry represented, on average, 256% more of total holdings than securities without industry. On the other hand, *Financials* had the lowest coefficient of 0.77, meaning that those stocks represented nonetheless, on average, more weight than no industry securities. The coefficients for industry times the Covid dummy show that, after the Covid-19 stock market crash, there was significant change in the allocation of holdings across industries. Except for Consumer Staples × Covid, all coefficients are statistically significant at 10% significance level. The industries that experienced the largest increases and decreases in holdings were, respectively, *Energy* and *Technology*, with respective covid coefficients of 0.25 and -0.36. In column (3), we have the coefficients estimated with the fixed effects method, which are the same as with pooled OLS, with the difference being that they are all statistically significant at 1% significance. In Table 5 it is shown the descriptive statistics for the coefficients of all regressions. On average, the coefficients for industry were 2.047 in absolute value, with a standard deviation of 0.508. For the coefficients of industry×Covid, the average absolute value was 0.201, with a standard deviation of 0.073. The R^2 is 25.88% and 5.01% for the OLS and FE estimators, respectively. The OLS regression without covid variables had an R^2 of 25.71%, implying that the industry×Covid variables account for 0.17 pp of the variation in logHolders, for this model.

In columns (2) and (4) of Table 4, it is described the results for the pooled OLS and Fixed Effects regressions, respectively, using the full model with control variables. The ranking of industries is different in this model compared to the previous. Prior to Covid, the most preferred industry was *Consumer Staples,* with an OLS coefficient of 1.11. The least preferred was still *Financials,* with an OLS coefficient of 0.27. The coefficients for industry are all statistically significant at any level, in the OLS method. The changes in industry allocation with Covid were also different from the no controls regression. The coefficient of *Basic Materials* × *Covid* (the second smallest in the no controls model) is not statistically significant in the OLS method. The Covid coefficients for *Industrials*, *Telecommunications* and *Utilities* are also not statistically significant in the OLS method, while *Consumer Staples* and *Technology* are only significant at the 5% level. Furthermore, the Covid coefficient for *Technology* was around a fourth in absolute value (-0.076). The model was not able to explain the changes in *Consumer Discretionary*, *Energy*, *Health Care*, *Real Estate*, and *Financials* as their Covid coefficients and respective z-scores did not change significantly in absolute value. However, it is important to note that, even for those changes that were not explained away, the results of the full model are more in line with what would be expected. The Covid coefficient for *Health Care* increased to a positive value of 0.16, implying a preference for that industry during Covid, which would be expected during the pandemic. The same was observed for *Consumer Staples*, another industry that should be more attractive in times of financial uncertainty, showing a positive coefficient of 0.13. The results of the FE estimator were similar, although *Health Care*, *Real Estate* and *Financials* showed much smaller Covid coefficients (0.085, -0.064, and -0.023, respectively) relative to their coefficients in the FE regression with no controls (-0.15, -0.14, and -0.16 respectively). Joining the findings of both regression methods, the factors we controlled for were unable to explain the changes in *Consumer Discretionary* and *Energy*, and explained only partly the changes in *Health Care*, *Real Estate*, *Consumer Staples*, and *Financials*. The R^2 of both methods were substantially larger than in the no controls regressions, being 62.55% and 31.65% for the OLS and FE, respectively.

In Table 5, it is shown the descriptive statistics for the coefficients of all regressions. As expected, for the full model, the coefficients of industries for before covid are on average closer to zero than in the regression with no controls, with an average absolute value 0.7407, for the OLS method. Their distribution is also more concentrated around the mean, with a standard deviation of 0.2311. The full model showed equally a decrease in the difference between the periods industry coefficients. For the OLS estimation, the average absolute value (a.a.v) of the industry×Covid coefficients decreased to 0.1196. The FE method also saw similar decreases, with an a.a.v of 0.1077. In the last row of Table 5 it is shown the difference between the R^2 of each regression and of the same regression without industry×Covid variables. The variation in industry×Covid variables accounts for 0.079 pp and 2.305 pp of the variation in holdings in the OLS and FE, respectively, which is slightly below 50% of the variation accounted for by industry×Covid in the model without controls.

In Table 6 to Table 9 it is described the results for the variable importance regressions, where each regression is the full model, excluding one of the control factors. In Table 10 and Table 11 it is shown the descriptive statistics of the coefficients of interest for each regression. Removing the trading volume resulted in the largest a.a.v and R^2 difference for the industry×Covid variables, for both OLS and FE. The R^2 difference was 0.116 pp (OLS) and 3.395 pp (FE), implying that trading volume explains around 22% of the variation in holdings explained by the variation in Industry×Covid in the model with no controls. The second most important variable was Price/Earnings, with R^2 differences of 0.082 pp (OLS) and 2.294 pp (FE), implying that it accounts for around 1% of the explanatory power of Industry×Covid on holdings. The results for price changes variables showed opposite results between the OLS and FE method. The R^2 difference for OLS was around 9% lower (0.063 pp) than in the full model, while for FE it was around 4% higher (2.524 pp) than in the full model. Therefore, it is inconclusive whether buying stocks on large price changes played a role in the reallocation of holdings across industries.

Removing the factor beta also showed a decrease in the R^2 difference (0.077 pp in the OLS) from the model with controls, implying it did not play a role in the industry reallocation of holdings. Since beta is constant across time, it is not possible to omit it in the FE method and, therefore, we only have the OLS results. This result implies that there might be an omitted variable problem with our model, which correlates with beta and Industry×Covid, and with the dependent variable.

7. Conclusion

The paper tried to observe how retail investors reallocated their investments across industries following the Covid-19 financial market crash, and what factors might explain this reallocation. We have seen that Robinhood users largely overweighted certain industries prior to Covid-19, such as *Technology* and *Consumer Discretionary*, and, after the financial crash, they showed mostly contrary behavior to that of the rest of the market. They consistently underinvested in the industries or sectors that performed well, such as *Health Care* and *Technology*, and, at the same time, tended to overinvest in the industries that performed poorly.

To test which factors explain this change in industry allocation, we tried to control for behaviors that have previously been associated with Robinhood users and retail investors in the literature. Barber et al. (2020) showed that Robinhood investors are mostly driven by attention-based buying and herding behavior, and that are susceptible to the display of information on the app. So, we included the 52 weeks trading volume rolling average as a proxy for the "attention-level" of a stock. We also included a proxy for the app's Top Movers List, and the price Earnings ratio, averaged over six months, since it is displayed in the app's statistics page of a security. Grinblatt and Keloharju (2001) and Barber and Odean (2008) showed that retail investors in the 90s had bought securities that had increased or decreased dramatically in price recently. So, we included the number of days each security experienced one-day returns in the percentile 90 and 10 over the previous 20 business days. Finally, we included the CAPM historical beta to control for investors risk-aversion levels, which could have changed after the stock market drop.

We joined the mentioned factors as independent variables in our regression, together with industry dummies and industry dummies interacted with a Covid dummy, to see how they relate to the dependent variable of the number of users holding each security (as a percentage of total holdings), and whether they reduce the difference in industry allocation before and after the market drop. The Covid-19 event day was chosen as February 20, 2020, as it was the day the Dow Jones dropped the most during the pandemic. Our model results are significant and explain most of the variance of our dependent variable. We concluded that these factors account for around 55% of the variation in industry allocation after the Covid-19 market drop. However, not all changes in industry could be explained. These factors could account for a large part of the changes in of *Basic Materials*, *Industrials*, *Telecommunications*, *Technology*, and *Utilities*, for a smaller part of the changes in *Health Care*, *Real Estate*, *Consumer Staples,* and *Financials*, but not for the changes in *Energy* and *Consumer Discretionary*. However, the full model showed an increase in the preference for the *Health Care* and *Consumer Staples* industries during Covid, where a decrease in holdings had been previously identified when not controlling for any factors. These results are more in line with what would be expected during the pandemic.

The most relevant variable to explain the change in industry allocation was a security's trading volume, accounting for around 22% of the variation. The second most relevant factor was the company's Price/Earnings, explaining around 1% of the variation in industry allocation. Hence, during the pandemic, Robinhood users were drawn to industries with highly liquid stocks and with high earnings compared to the stock's price. However, their preference for high Price/Earnings was lower during Covid than previously. This is in accordance with previous research, which concluded that Robinhood users acted, to a certain extent, as contrarian market forces, buying the stocks that institutional investors were selling in droves (Glossner et al. 2020).

My research paper offers a first attempt at identifying which factors lead to a change in retail investors' industry preference following a large financial market drop and an increase in uncertainty, as was the case for Covid-19. While most of this variation was explained, the variation in the *Energy* and *Consumer Discretionary* industries could not be accounted for in my model. This leaves a space for future research to explore. One thing that was noted and could not be properly accounted for due to the nature of the Robintrack data, was the surge in new users following the Covid event day. It is possible that these new users initially bought stocks differently from the rest of the Robinhood users, and thus have skewed the results; this could be part of the explanation for the change in these industries. Future research, using another dataset of retail investors transactions with user id, which is lacking in my dataset, could explore this possibility.

References

Baker, S. R., Bloom, N., Davis, S. J., Kost, K. J., Sammon, M. C., and Viratyosin, T. 2020. "The Unprecedented Stock Market Reaction to COVID-19." The Review of Asset Pricing Studies, Volume 10, Issue 4, December 2020, Pages 742–758

Barber, Brad M., and Odean Terrance. 2005. "All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." EFA 2005 Moscow Meetings Paper 51.

Barber, Brad M., Terrance Odean, and Ning Zhu. 2006. "Systematic Noise." AFA 2004 San Diego Meetings, EFA 2005 Moscow Meetings Paper.

Barber, Brad M., Terrance Odean, and Ning Zhu. 2008. "Do retail trades move markets?" In: The Review of Financial Studies 22.1, pp. 151–186.

Barber, Brad M., Xing Huang, Terrance Odean, and Christopher Schwarz. 2020. "Attention Induced Trading and Returns: Evidence from Robinhood Users." Journal of Finance, Forthcoming 73.

Burch, T.R., Emery, D.R., Fuerst, M.E. 2016. "Who moves markets in a sudden marketwide crisis? Evidence from 9/11." J. Financ. Quant. Anal. 51, 463–487.

Glossner, Simon, Pedro Matos, Stefano Ramelli, and Alexander F. Wagner (Sept. 2020). Where do institutional investors seek shelter when disaster strikes? Evidence from COVID-19. Tech. rep. Virginia/Darden and Zuric

Grinblatt, Mark and Matti Keloharju. 2000. "The investment behavior and performance of various investor types: a study of Finland's unique data set". In: Journal of Financial Economics 55.1, pp. 43–67.

Grömping, U. 2015. "Variable importance in regression models." Wiley Interdisciplinary Reviews: Computational Statistics, 7(2), 137–152. https://doi.org/10.1002/wics.1346

Lam, J. 2022. "Robinhood Review 2022: An Investing App to Avoid." Time. July 8. Accessed October 18, 2022. https://time.com/nextadvisor/investing/brokerage-reviews/robinhood-review/

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

Moss, Austin, James P. Naughton, and Clare Wang. 2020. "The Irrelevance of ESG Disclosure to Retail Investors: Evidence from Robinhood." Tech. rep. University of Iowa, University of Virgina, and University of Colorado.

Odean, T. 1999. "Do investors trade too much?" American Economic Review, 89, 1279–1298.

Ozik, Gideon, Ronnie Sadka, and Siyi Chen. Nov. 2020. "Flattening the Illiquidity Curve: Retail Trading During the COVID-19 Lockdown." Tech. rep. Edhec, Boston College, and Chinese University of Hong Kong.

Ponczek, Sarah. 2020. "Robintrack, Chronicler of Day Trader Stock Demand, To Shut." Bloomberg. August 8. Accessed October 18, 2022. https://www.bloomberg.com/news/articles/2020-08-08/robintrack-chronicler-of-day-trader-stock-demand-to-shut-down.

Popper, N. 2020. "Robinhood has lured young traders, sometimes with devastating results." The New York Times. July 8. Accessed October 18, 2022. https://www.nytimes.com /2020/07/08/technology/robinhood-risky-trading.html

"Robinhood raises $280 million in series F funding led by Sequoia. Under the Hood." 2020. July 13. Accessed October 18, 2022. https://blog.robinhood.com/news/2020/5/4/robinhood-raises-280-million-in-series-f-funding-led-by-sequoia

R. Ortmann, M. Pelster, S.T. Wengerek. 2020. "COVID-19 and investor behaviour." Finance Res. Lett., 37 (2020), Article 101717.

S. Baek, S.K. Mohanty, M. Glambosky. 2020. "COVID-19 and stock market volatility: An industry level analysis." Finance research letters, 37 (2020), Article 101748

"Viewing stock detail pages." n.d. Robinhood. Accessed October 18, 2022. https://robinhood .com /us/en/support/articles/viewing-stock-detail-pages

Wang, A.Y., Young, M. 2020. "Terrorist attacks and investor risk preference: Evidence from mutual fund flows." Journal of Financial Economics 137 (2), 491–514.

Welch, Ivo. 2020. "The Wisdom of the Robinhood Crowd." NBER Working Paper No. w27866 52.

Wooldridge, Jeffrey M. 2012. "Advanced Panel Data Methods." In Introductory Econometrics: A Modern Approach, Fifth Edition, 5th ed., 484–92. Mason OH: South-Western Cengage Learning.

APPENDIX A

Table 1: Descriptive statistics

This table reports the Descriptive statistics for all the model variables for the entirety of the regression sample, from 14th of August 2019 to 13th of August 2020.

	count	mean	std	min	25%	50%	75%	max
log Price Earnings	1244022	-3.84	3.49	-6.91	-6.91	-6.91	-0.20	7.01
Top List	1244022	0.05	0.30	0.00	0.00	0.00	0.00	9.00
log Trading Volume Trail	1244022	-1.92	2.15	-9.40	-3.41	-1.74	-0.39	5.03
Top Absolute Return	1244022	4.00	4.15	0.00	0.00	3.00	7.00	20.00
Beta	1244022	0.99	0.85	-8.27	0.56	0.97	1.36	11.09
log Holders	1244022	-6.24	2.02	-12.69	-7.65	-6.26	-4.91	1.72

Table 2: Variable correlation

This table reports the correlations for the non-categorical variables for the entirety of the regression sample, from 14th of August 2019 to 13th of August 2020.

	log Price Earnings	log Trading Volume	Top List	Top Absolute Return	Beta	log Holders
log Price Earnings	1	-	-	-	-	-
log Trading Volume	0.24	1	-	-	-	-
Top List	-0.03	0.13	1	-	-	-
Top Absolute Return	-0.08	0.10	0.26	1	-	-
Beta	0.03	0.19	0.08	0.15	1	-
log Holders	0.11	0.72	0.18	0.35	0.23	1

Table 3: Variance inflation factor

This table reports the Variance inflation factor for all the model variables for the entirety of the regression sample, from 14th of August 2019 to 13th of August 2020.

	VIF		VIF		VIF
Basic Materials	2.24	Utilities	2.04	Telecommunications × Covid	2.05
Consumer Discretionary	2.59	Basic Materials × Covid	2.09	Utilities × Covid	2.02
Consumer Staples	2.12	Consumer Discretionary × Covid	2.32	log Price Earnings	4.52
Energy	2.34	Consumer Staples × Covid	2.05	log Price Earnings × Covid	4.27
Financials	2.54	Energy × Covid	2.17	log Trading Volume Trail	3.64
Health Care	3.14	Financials × Covid	2.49	log Trading Volume Trail × Covid	3.56
Industrials	2.61	Health Care × Covid	2.59	Top List	1.13
Real Estate	2.12	Industrials × Covid	2.44	Top Absolute Return	3.27
Technology	2.54	Real Estate × Covid	2.09	Beta	4.99
Telecommunications	2.09	Technology × Covid	2.36	Beta × Covid	4.99

Table 4: Regression results

This table reports the estimated coefficients, for the model with and without controls, presented in section 3. Using Pooled OLS estimation method as presented in Equation (1), columns (1) and (2) coefficients are estimated without controls and with, respectively. Using the Fixed Effects estimation method as presented in Equation (2), columns (3) and (4) coefficients are estimated without and with controls, respectively. t-statistics calculated using the Newey-West method, with 21 lags, are shown in parentheses. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

log Holders	(1)	(2)	(3)	(4)
Intercept	-7.548*** (-526.773)	-6.245*** (-326.369)	0.000*** (0.000)	0.000*** (0.000)
Basic Materials	2.222*** (41.287)	0.617*** (15.198)	0.000*** (0.000)	0.000*** (0.000)
Basic Materials × Covid	-0.236*** (-3.298)	0.010 (0.190)	-0.236*** (-20.388)	-0.013 (-1.182)
Consumer Discretionary	2.398*** (67.040)	0.972*** (32.210)	0.000*** (0.000)	0.000*** (0.000)
Consumer Discretionary × Covid	0.137*** (2.898)	0.198*** (5.612)	0.137*** (10.888)	0.267*** (24.497)
Consumer Staples	2.211*** (39.686)	1.113*** (27.500)	0.000*** (0.000)	0.000*** (0.000)
Consumer Staples × Covid	-0.101 (-1.369)	0.129** (2.586)	-0.101*** (-5.809)	0.114*** (6.582)
Energy	2.507*** (54.682)	0.721*** (16.587)	0.000*** (0.000)	0.000*** (0.000)
Energy × Covid	0.250***	0.266***	0.250***	0.357***

	(3.972)	(4.913)	(14.481)	(22.898)
Financials	0.767*** (25.546)	0.271*** (10.984)	0.000*** (0.000)	0.000*** (0.000)
Financials × Covid	-0.163*** (-4.332)	-0.122*** (-4.312)	-0.163*** (-24.093)	-0.023*** (-3.313)
Health Care	2.490*** (84.009)	0.724*** (27.102)	0.000*** (0.000)	0.000*** (0.000)
Health Care × Covid	-0.148*** (-3.977)	0.156*** (4.929)	-0.148*** (-14.910)	0.085*** (9.349)
Industrials	1.508*** (49.408)	0.422*** (15.457)	0.000*** (0.000)	0.000*** (0.000)
Industrials × Covid	-0.174*** (-4.540)	-0.002 (-0.069)	-0.174*** (-21.414)	0.019** (2.269)
Real Estate	1.820*** (34.401)	0.830*** (19.502)	0.000*** (0.000)	0.000*** (0.000)
Real Estate × Covid	-0.138* (-1.952)	-0.189*** (-3.494)	-0.138*** (-10.478)	-0.064*** (-5.294)
Technology	2.557*** (63.731)	0.946*** (30.954)	0.000*** (0.000)	0.000*** (0.000)
Technology × Covid	-0.361*** (-6.922)	-0.076** (-2.043)	-0.361*** (-44.523)	-0.105*** (-12.463)
Telecommunications	2.069*** (26.749)	0.711*** (13.948)	0.000*** (0.000)	0.000*** (0.000)
Telecommunications × Covid	-0.219** (-2.079)	0.068 (1.016)	-0.219*** (-13.994)	0.044*** (2.894)
Utilities	1.969*** (34.772)	0.822*** (17.307)	0.000*** (0.000)	0.000*** (0.000)
Utilities × Covid	-0.279*** (-3.677)	-0.099 (-1.621)	-0.279*** (-22.414)	-0.095*** (-7.538)
log Trading Volume Trail	0.000*** (0.000)	0.569*** (154.609)	0.000*** (0.000)	2.676*** (67.377)
log Trading Volume Trail × Covid	0.000*** (0.000)	0.061*** (12.616)	0.000*** (0.000)	0.058*** (49.508)
log Price Earnings	0.000*** (0.000)	-0.069*** (-28.337)	0.000*** (0.000)	-0.025*** (-19.168)
log Price Earnings × Covid	0.000*** (0.000)	0.023*** (8.135)	0.000*** (0.000)	0.024*** (35.385)
Top List	0.000*** (0.000)	0.136*** (8.766)	0.000*** (0.000)	0.167*** (25.580)
Top Absolute Return	0.000*** (0.000)	0.085*** (58.827)	0.000*** (0.000)	0.028*** (43.516)
Beta	0.000*** (0.000)	0.121*** (14.077)	0.000*** (0.000)	0.000*** (0.000)
Beta × Covid	0.000*** (0.000)	-0.020* (-1.666)	0.000*** (0.000)	-0.021*** (-6.492)
R^2	25.88%	62.55%	5.01%	31.65%
Observations	1244022	1244022	1244022	1244022

Table 5: Regression coefficients statistics

This table reports the statistics of the industry coefficients, for the model with and without controls, presented in section 3, and the difference in R^2 between each model and the same model excluding the industry×Covid variables. Using Pooled OLS estimation method as presented in

Equation (1), columns (1) and (2) coefficients are estimated without controls and with, respectively. Using the Fixed Effects estimation method as presented in Equation (2), columns (3) and (4) coefficients are estimated without and with controls, respectively.

	(1)	(2)	(3)	(4)
Average |Industry Coefficient|	2.0471	0.7407	-	-
Standard Deviation Industry Coefficient	0.5077	0.2311	-	-
Average |t-statistic of industry|	47.3919	20.6135	-	-
Average |Industry × Covid Coefficient|	0.2005	0.1196	0.2005	0.1077
Standard Deviation Industry × Covid Coefficient	0.0728	0.0767	0.0728	0.1038
Average |t-statistic of industry × Covid|	3.5469	2.7986	18.4902	8.9345
ΔR2	0.1745%	0.0792%	5.0080%	2.3051%

Table 6: OLS regression results for variable importance

This table reports the estimated coefficients, for the model with controls, presented in section 3, with one control factor removed, using the OLS method. Column (1) removes Top Absolute Return and Top List, column (2) removes log Price Earnings Ratio and log Price Earnings Ratio × Covid, column (3) removes log Trading Volume Trail and log Trading Volume Trail × Covid, and column (4) removes Beta and Beta × Covid. t-statistics calculated using the Newey-West method, with 21 lags, are shown in parentheses. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

log Holders	(1)	(2)	(3)	(4)
Intercept	-6.460*** (-325.475)	-5.895*** (-393.461)	-8.020*** (-323.325)	-6.168*** (-331.793)
Basic Materials	1.310*** (31.662)	0.310*** (7.670)	1.906*** (33.771)	0.654*** (16.591)
Basic Materials × Covid	-0.091* (-1.705)	0.023 (0.451)	-0.045 (-0.618)	-0.006 (-0.134)
Consumer Discretionary	1.582*** (55.075)	0.621*** (22.156)	2.134*** (50.413)	1.029*** (34.859)
Consumer Discretionary × Covid	0.228*** (6.273)	0.216*** (6.100)	0.223*** (4.451)	0.178*** (5.293)
Consumer Staples	1.671*** (40.850)	0.777*** (19.961)	2.125*** (35.497)	1.121*** (27.919)
Consumer Staples × Covid	0.058 (1.114)	0.157*** (3.116)	0.036 (0.476)	0.118** (2.389)
Energy	1.439*** (33.607)	0.397*** (9.554)	2.076*** (39.888)	0.797*** (18.771)
Energy × Covid	0.259*** (4.728)	0.299*** (5.578)	0.329*** (4.987)	0.241*** (4.611)
Financials	0.600***	-0.078***	0.786***	0.312***

	(24.266)	(-3.569)	(22.315)	(13.008)
Financials × Covid	0.010 (0.347)	-0.103*** (-3.663)	-0.150*** (-3.734)	-0.139*** (-5.293)
Health Care	1.467*** (58.967)	0.509*** (19.767)	1.842*** (51.308)	0.804*** (30.878)
Health Care × Covid	-0.004 (-0.105)	0.145*** (4.596)	0.193*** (4.569)	0.135*** (4.560)
Industrials	0.936*** (34.015)	0.058** (2.363)	1.294*** (34.465)	0.500*** (19.100)
Industrials × Covid	-0.007 (-0.218)	0.036 (1.115)	-0.057 (-1.313)	-0.026 (-0.893)
Real Estate	1.242*** (28.377)	0.455*** (11.005)	1.833*** (32.516)	0.868*** (20.513)
Real Estate × Covid	-0.070 (-1.237)	-0.167*** (-3.050)	-0.149** (-2.068)	-0.207*** (-3.876)
Technology	1.521*** (50.339)	0.660*** (22.609)	2.147*** (47.905)	1.037*** (35.192)
Technology × Covid	-0.156*** (-3.967)	-0.069* (-1.841)	-0.130** (-2.315)	-0.100*** (-2.893)
Telecommunications	1.234*** (24.074)	0.436*** (8.540)	1.840*** (24.033)	0.747*** (14.647)
Telecommunications × Covid	-0.013 (-0.196)	0.075 (1.112)	-0.025 (-0.244)	0.053 (0.797)
Utilities	1.268*** (25.682)	0.488*** (10.738)	2.045*** (34.714)	0.800*** (16.799)
Utilities × Covid	-0.123* (-1.950)	-0.098 (-1.616)	-0.170** (-2.248)	-0.105* (-1.721)
log Trading Volume Trail	0.559*** (149.717)	0.554*** (159.036)	0.000*** (0.000)	0.575*** (156.912)
log Trading Volume Trail × Covid	0.060*** (12.185)	0.083*** (20.085)	0.000*** (0.000)	0.060*** (12.465)
log Price Earnings	-0.110*** (-45.053)	0.000*** (0.000)	-0.049*** (-14.790)	-0.072*** (-30.131)
log Price Earnings × Covid	0.023*** (7.668)	0.000*** (0.000)	0.033*** (9.688)	0.024*** (9.039)
Top List	0.000*** (0.000)	0.135*** (8.816)	0.586*** (28.326)	0.144*** (9.486)
Top Absolute Return	0.000*** (0.000)	0.102*** (75.138)	0.043*** (19.862)	0.085*** (58.703)
Beta	0.126*** (13.138)	0.148*** (16.879)	0.302*** (21.443)	0.000*** (0.000)
Beta × Covid	-0.033** (-2.410)	-0.048*** (-4.065)	-0.045** (-2.367)	0.000*** (0.000)
R^2	60.62%	62.01%	29.50%	62.36%
Observations	1244022	1244022	1244022	1244022

Table 7: FE regression results for variable importance

This table reports the estimated coefficients, for the model with controls, presented in section 3, with one control factor removed, using the FE method. Column (1) removes Top Absolute Return and Top List, column (2) removes log Price Earnings Ratio and log Price Earnings Ratio ×

Covid, and column (3) removes log Trading Volume Trail and log Trading Volume Trail × Covid. t-statistics calculated using the Newey-West method, with 21 lags, are shown in parentheses. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

log Holders	(1)	(2)	(3)
Intercept	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Basic Materials	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Basic Materials × Covid	-0.041*** (-3.663)	-0.028** (-2.453)	-0.050*** (-4.232)
Consumer Discretionary	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Consumer Discretionary × Covid	0.283*** (24.223)	0.278*** (25.544)	0.245*** (20.015)
Consumer Staples	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Consumer Staples × Covid	0.094*** (5.310)	0.120*** (6.862)	0.030* (1.744)
Energy	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Energy × Covid	0.377*** (22.987)	0.351*** (22.617)	0.388*** (22.797)
Financials	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Financials × Covid	0.023*** (3.216)	0.018*** (2.608)	-0.123*** (-16.881)
Health Care	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Health Care × Covid	0.023** (2.511)	0.044*** (4.838)	0.157*** (14.184)
Industrials	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Industrials × Covid	0.020** (2.334)	0.051*** (6.398)	-0.052*** (-5.648)
Real Estate	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Real Estate × Covid	-0.017 (-1.340)	-0.031** (-2.573)	-0.113*** (-8.883)
Technology	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Technology × Covid	-0.132*** (-15.385)	-0.111*** (-13.161)	-0.148*** (-15.333)
Telecommunications	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Telecommunications × Covid	0.020 (1.317)	0.031** (2.046)	-0.037** (-2.328)
Utilities	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Utilities × Covid	-0.106*** (-8.311)	-0.094*** (-7.551)	-0.188*** (-14.534)
log Trading Volume Trail	2.831*** (67.208)	2.590*** (65.285)	0.000*** (0.000)
log Trading Volume Trail × Covid	0.058*** (47.605)	0.079*** (78.755)	0.000*** (0.000)

log Price Earnings	-0.027*** (-19.522)	0.000*** (0.000)	-0.033*** (-22.058)
log Price Earnings × Covid	0.024*** (35.000)	0.000*** (0.000)	0.034*** (55.400)
Top List	0.000*** (0.000)	0.165*** (25.331)	0.198*** (24.683)
Top Absolute Return	0.000*** (0.000)	0.028*** (44.611)	0.032*** (43.054)
Beta	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Beta × Covid	-0.024*** (-7.050)	-0.055*** (-18.013)	-0.046*** (-12.177)
R^2	27.41%	30.31%	16.57%
Observations	1244022	1244022	1244022

Table 8: OLS regression results for variable importance (no industry×Covid)

This table reports the estimated coefficients, for the model with controls, presented in section 3, without industry×Covid variable and with one control factor removed, using the OLS method. Column (1) removes Top Absolute Return and Top List, column (2) removes log Price Earnings Ratio and log Price Earnings Ratio × Covid, column (3) removes log Trading Volume Trail and log Trading Volume Trail × Covid, and column (4) removes Beta and Beta × Covid. t-statistics calculated using the Newey-West method, with 21 lags, are shown in parentheses. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

log Holders	(1)	(2)	(3)	(4)
Intercept	-6.463*** (-325.540)	-5.894*** (-393.321)	-8.026*** (-323.693)	-6.173*** (-331.565)
Basic Materials	1.266*** (41.243)	0.327*** (10.989)	1.895*** (44.105)	0.662*** (21.645)
Consumer Discretionary	1.698*** (75.376)	0.733*** (33.395)	2.258*** (65.367)	1.128*** (46.431)
Consumer Staples	1.702*** (56.488)	0.858*** (30.321)	2.154*** (47.846)	1.189*** (39.234)
Energy	1.570*** (50.225)	0.553*** (17.927)	2.253*** (54.561)	0.927*** (28.441)
Financials	0.607*** (30.075)	-0.128*** (-7.612)	0.720*** (23.957)	0.250*** (12.260)
Health Care	1.466*** (74.078)	0.588*** (28.389)	1.949*** (67.876)	0.881*** (40.883)
Industrials	0.935*** (42.659)	0.079*** (4.141)	1.277*** (41.065)	0.496*** (22.390)
Real Estate	1.210*** (37.351)	0.374*** (12.321)	1.769*** (40.964)	0.772*** (23.962)
Technology	1.445*** (62.344)	0.629*** (28.340)	2.093*** (60.877)	0.996*** (42.319)
Telecommunications	1.229*** (32.964)	0.477*** (13.057)	1.836*** (32.330)	0.781*** (20.933)

Utilities	1.208*** (33.392)	0.441*** (13.063)	1.969*** (43.602)	0.756*** (21.209)
log Trading Volume Trail	0.559*** (154.471)	0.553*** (161.983)	0.000*** (0.000)	0.571*** (162.173)
log Trading Volume Trail × Covid	0.060*** (13.192)	0.085*** (21.730)	0.000*** (0.000)	0.068*** (15.561)
log Price Earnings	-0.111*** (-46.136)	0.000*** (0.000)	-0.049*** (-14.666)	-0.070*** (-29.597)
log Price Earnings × Covid	0.023*** (8.164)	0.000*** (0.000)	0.030*** (8.968)	0.018*** (7.417)
Top List	0.000*** (0.000)	0.139*** (9.058)	0.590*** (28.386)	0.147*** (9.663)
Top Absolute Return	0.000*** (0.000)	0.101*** (74.743)	0.042*** (19.340)	0.083*** (58.058)
Beta	0.121*** (14.099)	0.134*** (17.332)	0.298*** (23.933)	0.000*** (0.000)
Beta × Covid	-0.022** (-2.272)	-0.020** (-2.307)	-0.039*** (-2.830)	0.000*** (0.000)
R^2	60.55%	61.93%	29.38%	62.28%
Observations	1244022	1244022	1244022	1244022

Table 9: FE regression results for variable importance (no industry×Covid)

This table reports the estimated coefficients, for the model with controls, presented in section 3, without industry×Covid variable and with one control factor removed, using the FE method. Column (1) removes Top Absolute Return and Top List, column (2) removes log Price Earnings Ratio and log Price Earnings Ratio × Covid, and column (3) removes log Trading Volume Trail and log Trading Volume Trail × Covid. t-statistics calculated using the Newey-West method, with 21 lags, are shown in parentheses. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

log Holders	(1)	(2)	(3)
Intercept	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Basic Materials	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Consumer Discretionary	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Consumer Staples	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Energy	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Financials	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)

Health Care	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Industrials	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Real Estate	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Technology	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Telecommunications	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Utilities	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
log Trading Volume Trail	2.867*** (68.228)	2.620*** (66.635)	0.000*** (0.000)
log Trading Volume Trail × Covid	0.057*** (49.755)	0.078*** (78.748)	0.000*** (0.000)
log Price Earnings	-0.032*** (-21.657)	0.000*** (0.000)	-0.036*** (-22.849)
log Price Earnings × Covid	0.024*** (36.498)	0.000*** (0.000)	0.032*** (50.154)
Top List	0.000*** (0.000)	0.171*** (25.615)	0.205*** (24.652)
Top Absolute Return	0.000*** (0.000)	0.029*** (45.001)	0.030*** (39.431)
Beta	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Beta × Covid	0.002 (0.769)	-0.024*** (-10.289)	-0.037*** (-13.639)
R^2	0.2488%	0.2802%	0.1317%
Observations	1244022	1244022	1244022

Table 10: OLS regression coefficients' statistics for variable importance

This table reports the statistics of the industry coefficients, for the model with controls, presented in section 3, with one control factor removed, using the OLS method, and the difference in R^2 between each model and the same model excluding the industry×Covid variables. Column (1) removes Top Absolute Return and Top List, column (2) removes log Price Earnings Ratio and log Price Earnings Ratio × Covid, column (3) removes log Trading Volume Trail and log Trading Volume Trail × Covid, and column (4) removes Beta and Beta × Covid.

	(1)	(2)	(3)	(4)
Average |Industry Coefficient|	1.2972	0.4353	1.8207	0.7881
Standard Deviation Industry Coefficient	0.2930	0.2125	0.4011	0.2269
Average |z-score	36.9922	12.5393	36.9841	23.1686

of industry|				
Average |Industry * Covid Coefficient|	0.0926	0.1262	0.1369	0.1189
Standard Deviation Industry * Covid Coefficient	0.0859	0.0780	0.0887	0.0695
Average |z-score of industry*covid|	1.9855	2.9307	2.4566	2.9509
ΔR2	0.0634 p.p.	0.0819 p.p.	0.1164 p.p.	0.0770 p.p.

Table 11: FE regression coefficients' statistics for variable importance

This table reports the statistics of the industry coefficients, for the model with controls, presented in section 3, with one control factor removed, using the FE method, and the difference in R2 between each model and the same model excluding the industry×Covid variables. Column (1) removes Top Absolute Return and Top List, column (2) removes log Price Earnings Ratio and log Price Earnings Ratio × Covid, and column (3) removes log Trading Volume Trail and log Trading Volume Trail × Covid.

	(1)	(2)	(3)
Average |Industry Coefficient|	-	-	-
Standard Deviation Industry Coefficient	-	-	-
Average |t-statistic of industry|	-	-	-
Average |Industry * Covid Coefficient|	0.1033	0.1051	0.1390
Standard Deviation Industry * Covid Coefficient	0.1153	0.1054	0.1022
Average |t-statistic of industry*covid|	8.2361	8.7865	11.5072
ΔR2	2.5239 p.p.	2.2944 p.p.	3.3953 p.p.

Table 12: Regression results (no industry×Covid)

This table reports the estimated coefficients, for the model with and without controls, presented in section 3, without the industry×Covid variable. Using Pooled OLS estimation method as presented in Equation (1), without the industry×Covid variable, columns (1) and (2) coefficients are estimated without controls and with, respectively. Using the Fixed Effects estimation method as presented in Equation (2), without the industry×Covid variable, column (3) coefficients are estimated with controls. t-statistics calculated using the Newey-West method, with 21 lags, are shown in parentheses. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

log Holders	(1)	(2)	(3)
Intercept	-7.548*** (-526.773)	-6.250*** (-326.121)	0.000*** (0.000)
Basic Materials	2.104*** (52.976)	0.632*** (20.350)	0.000*** (0.000)
Consumer Discretionary	2.466*** (87.016)	1.081*** (44.148)	0.000*** (0.000)
Consumer Staples	2.160*** (53.076)	1.186*** (39.011)	0.000*** (0.000)
Energy	2.632*** (74.034)	0.864*** (26.261)	0.000*** (0.000)
Financials	0.685*** (28.348)	0.218*** (10.600)	0.000*** (0.000)
Health Care	2.416*** (100.823)	0.812*** (37.318)	0.000*** (0.000)
Industrials	1.421*** (58.207)	0.430*** (19.133)	0.000*** (0.000)
Real Estate	1.751*** (44.532)	0.744*** (23.115)	0.000*** (0.000)
Technology	2.377*** (77.702)	0.917*** (38.323)	0.000*** (0.000)
Telecommunications	1.960*** (34.812)	0.753*** (20.211)	0.000*** (0.000)
Utilities	1.830*** (43.689)	0.780*** (22.022)	0.000*** (0.000)
log Trading Volume Trail	0.000*** (0.000)	0.566*** (159.003)	2.713*** (68.716)
log Trading Volume Trail × Covid	0.000*** (0.000)	0.068*** (15.393)	0.060*** (53.678)
log Price Earnings	0.000*** (0.000)	-0.067*** (-27.857)	-0.028*** (-20.469)
log Price Earnings × Covid	0.000*** (0.000)	0.018*** (6.656)	0.022*** (33.589)
Top List	0.000*** (0.000)	0.139*** (8.946)	0.173*** (25.761)
Top Absolute Return	0.000*** (0.000)	0.084*** (58.143)	0.028*** (42.089)
Beta	0.000*** (0.000)	0.114*** (14.755)	0.000*** (0.000)
Beta × Covid	0.000*** (0.000)	-0.006 (-0.617)	0.005** (2.040)
R2	25.71%	62.47%	29.34%
Observations	1244022	1244022	1244022

APPENDIX B

Figure 1: Missing values in Robintrack data

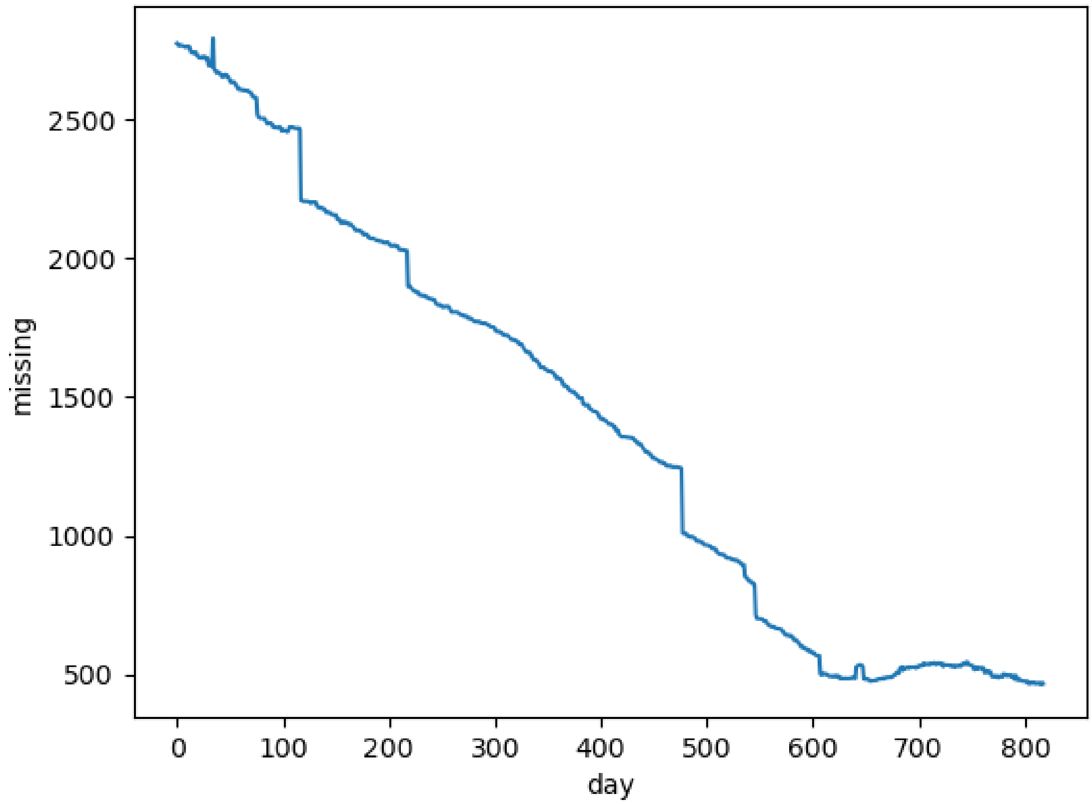


Figure 1 describes the number of missing data in the Robintrack dataset, for the entire 802 day-observations

Figure 2: Quote type distribution

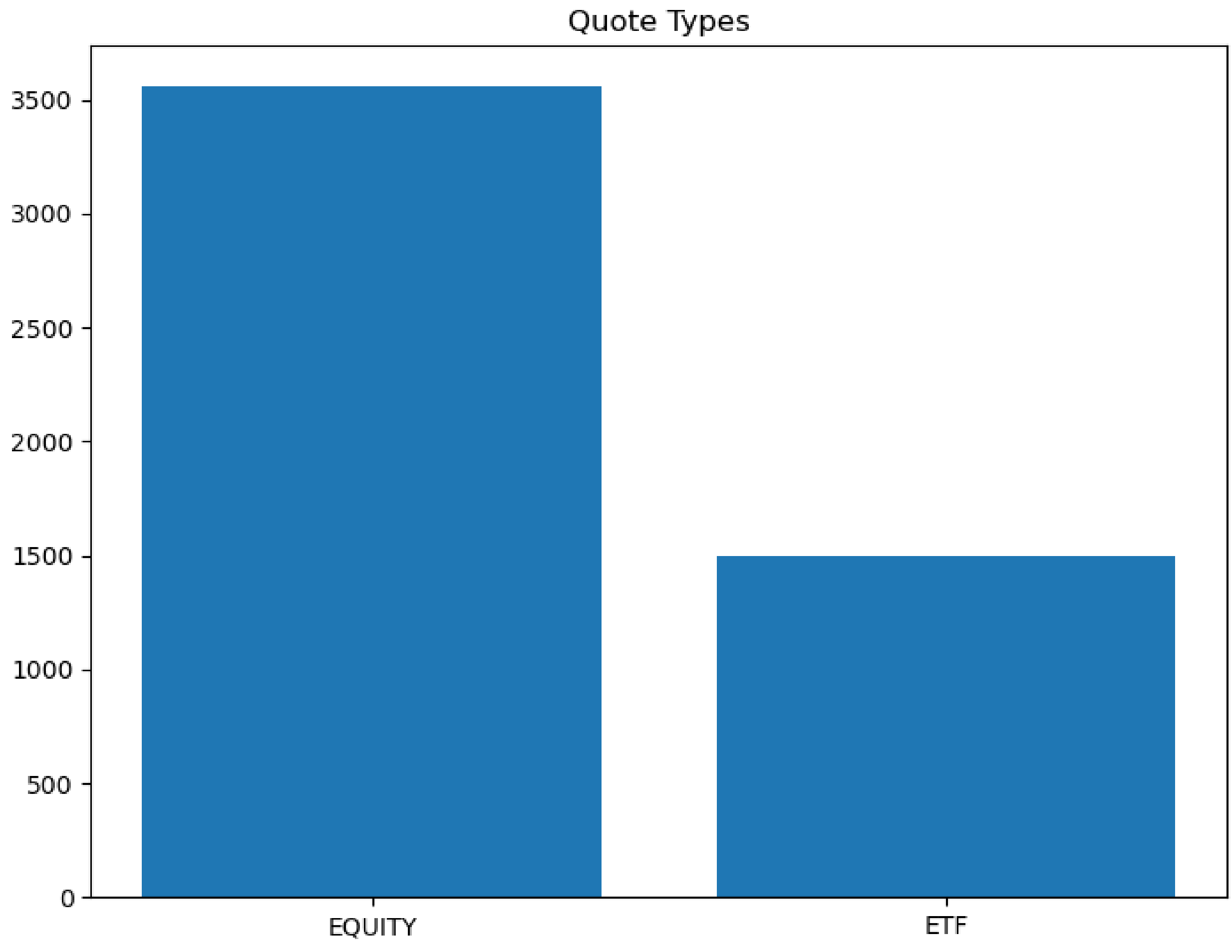


Figure 2 shows the quote type distribution of stocks in the final dataset, after all data transformation and cleaning.

Figure 3: Industry distribution

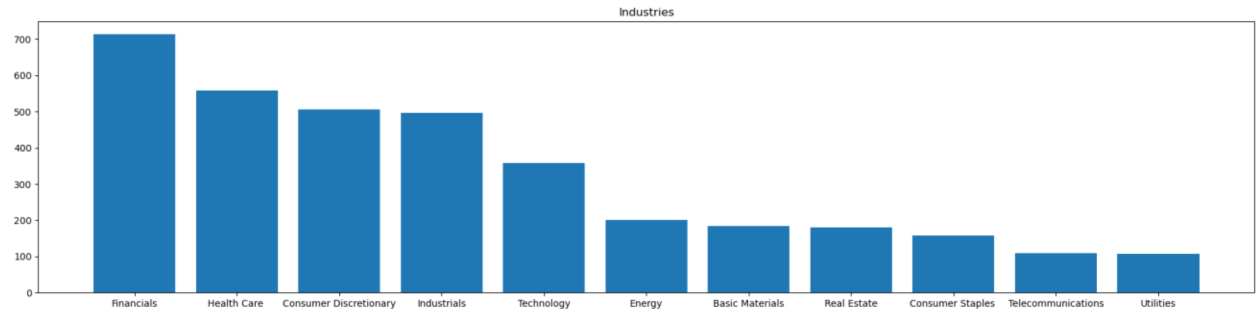


Figure 3 shows the industry distribution of stocks in the final dataset, after all data transformation and cleaning.

Figure 4: QQ Plot

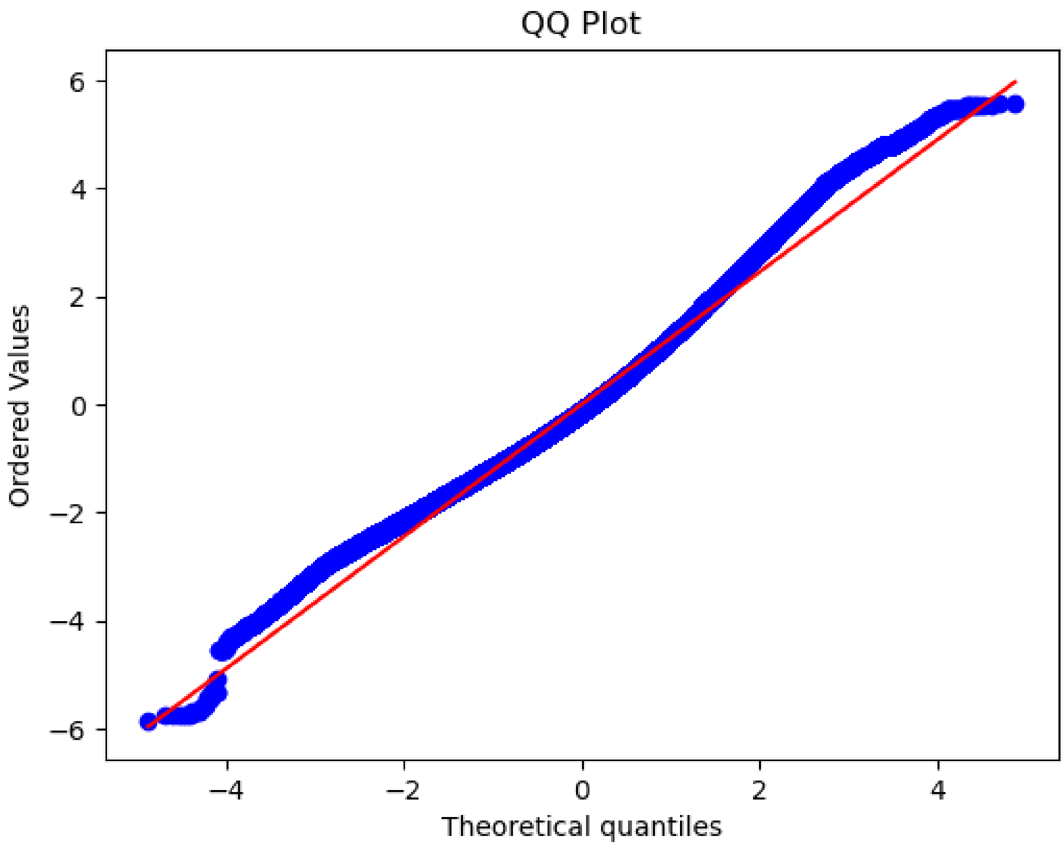


Figure 4 shows the QQ plot for testing the normality of errors in the final dataset used for the regression.

Figure 5: Homoskedasticity

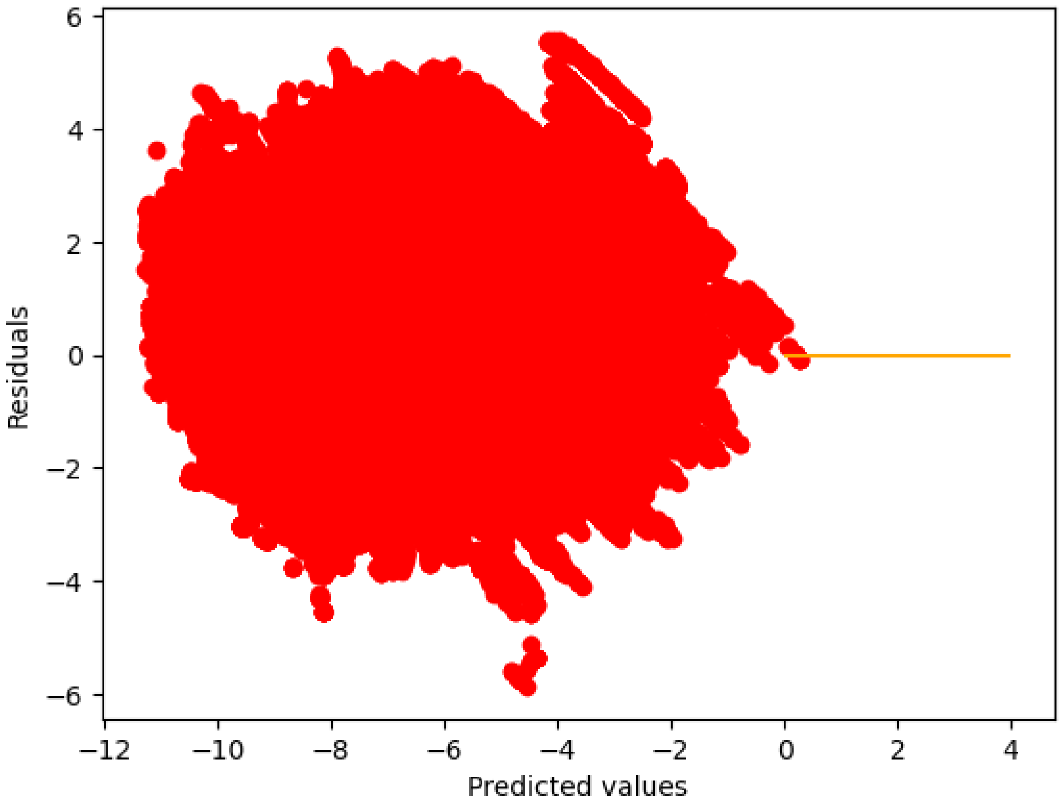


Figure 5 shows the homoscedasticity plot in the final dataset used for the regression.

Figure 6: Serial correlation of error terms

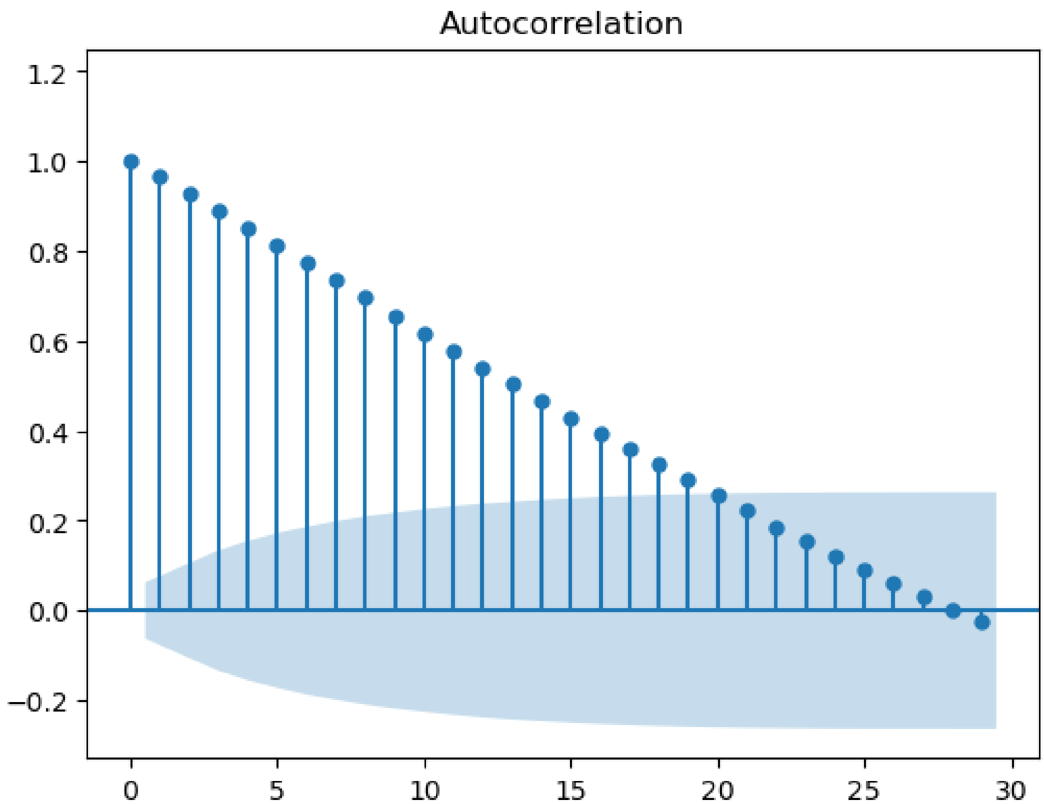


Figure 6 shows the plot for the autocorrelation of errors in the final dataset used for the regression, with the y-axis being the correlation and the x-axis the number of lags. Points above or below the blue area are statistically different from zero at the 5% significance level.

Figure 7: S&P 500 index and aggregate RH holdings

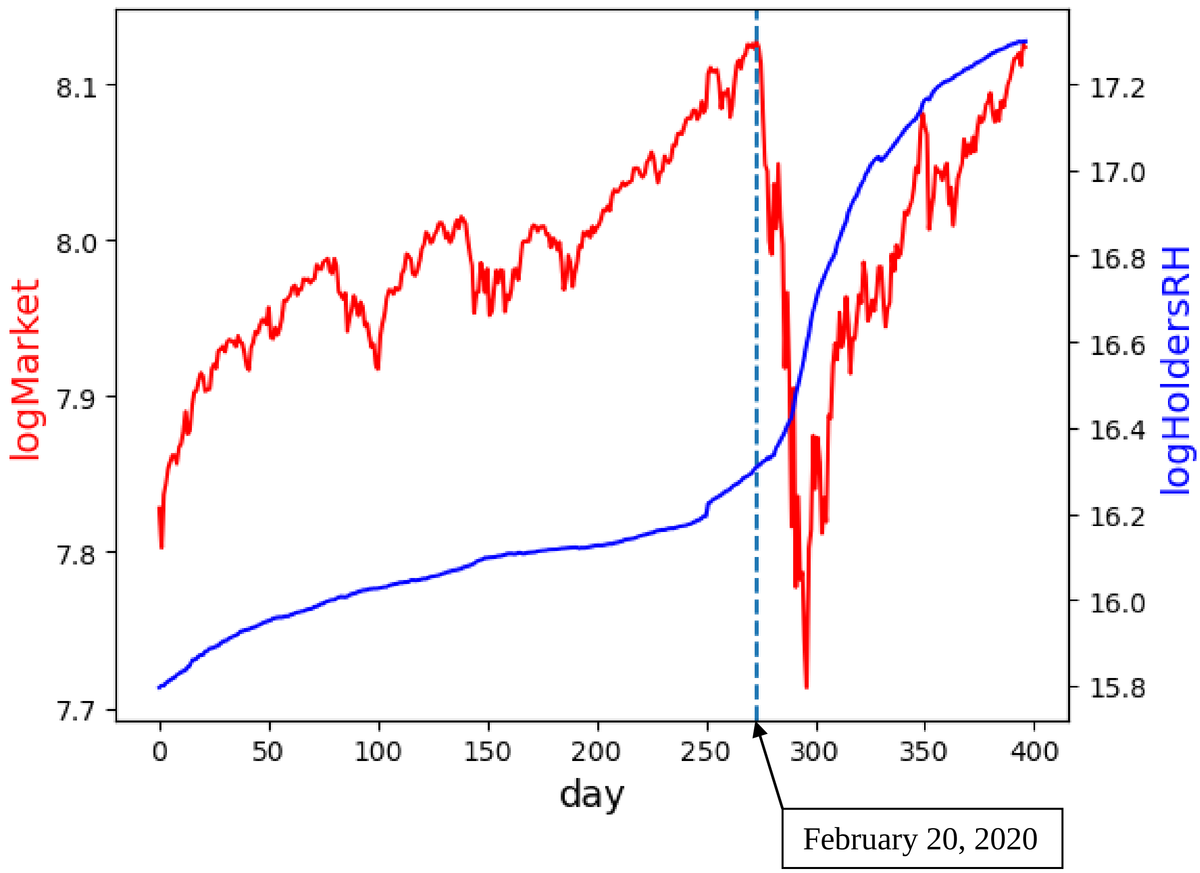


Figure 7 presents a time trend for the S&P 500 index and the sum of all Robinhood holdings in the final dataset, both log-transformed, in red and blue, respectively.

Figure 8: Aggregate holdings per industry, relative to an equal weighted portfolio

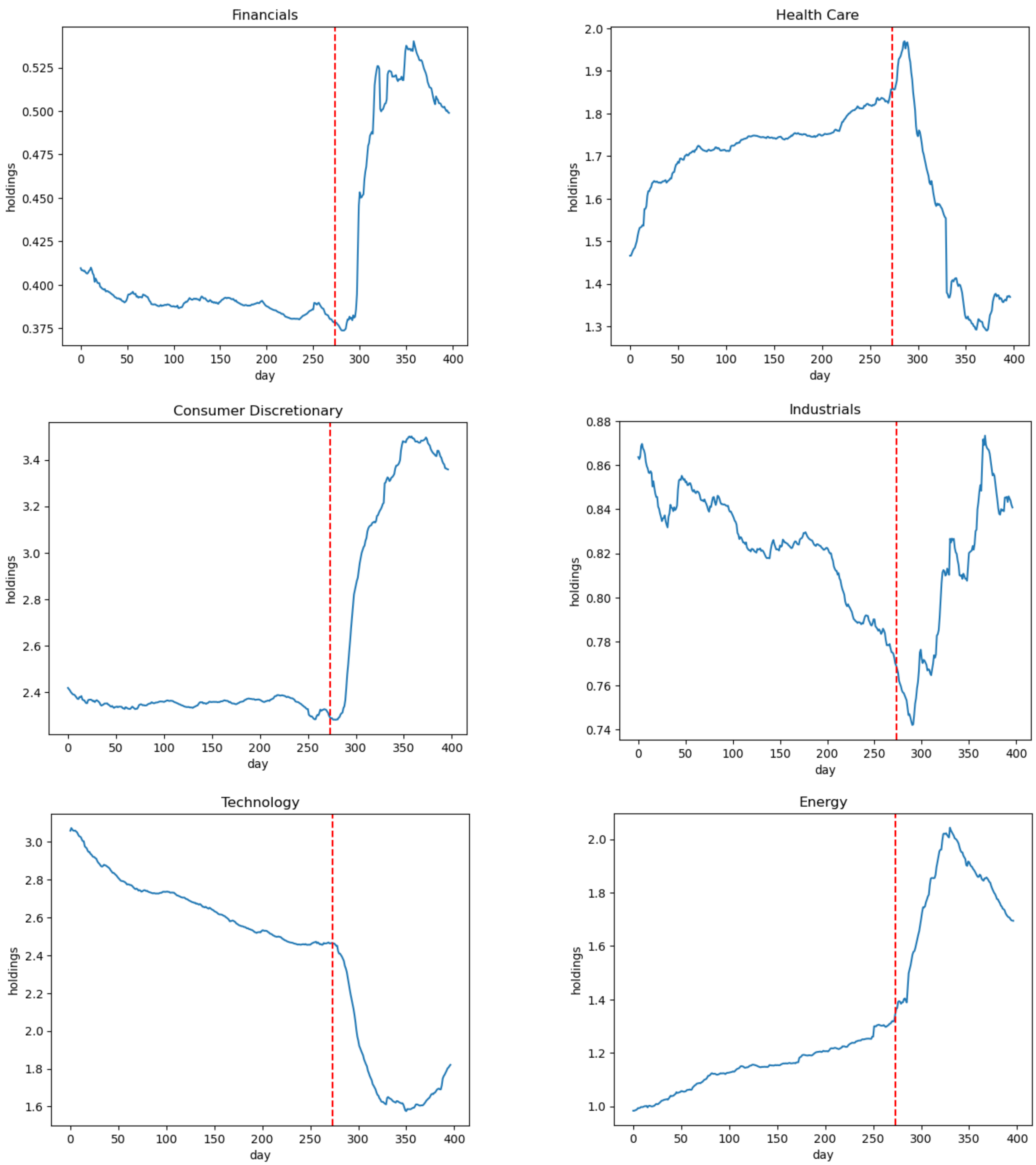
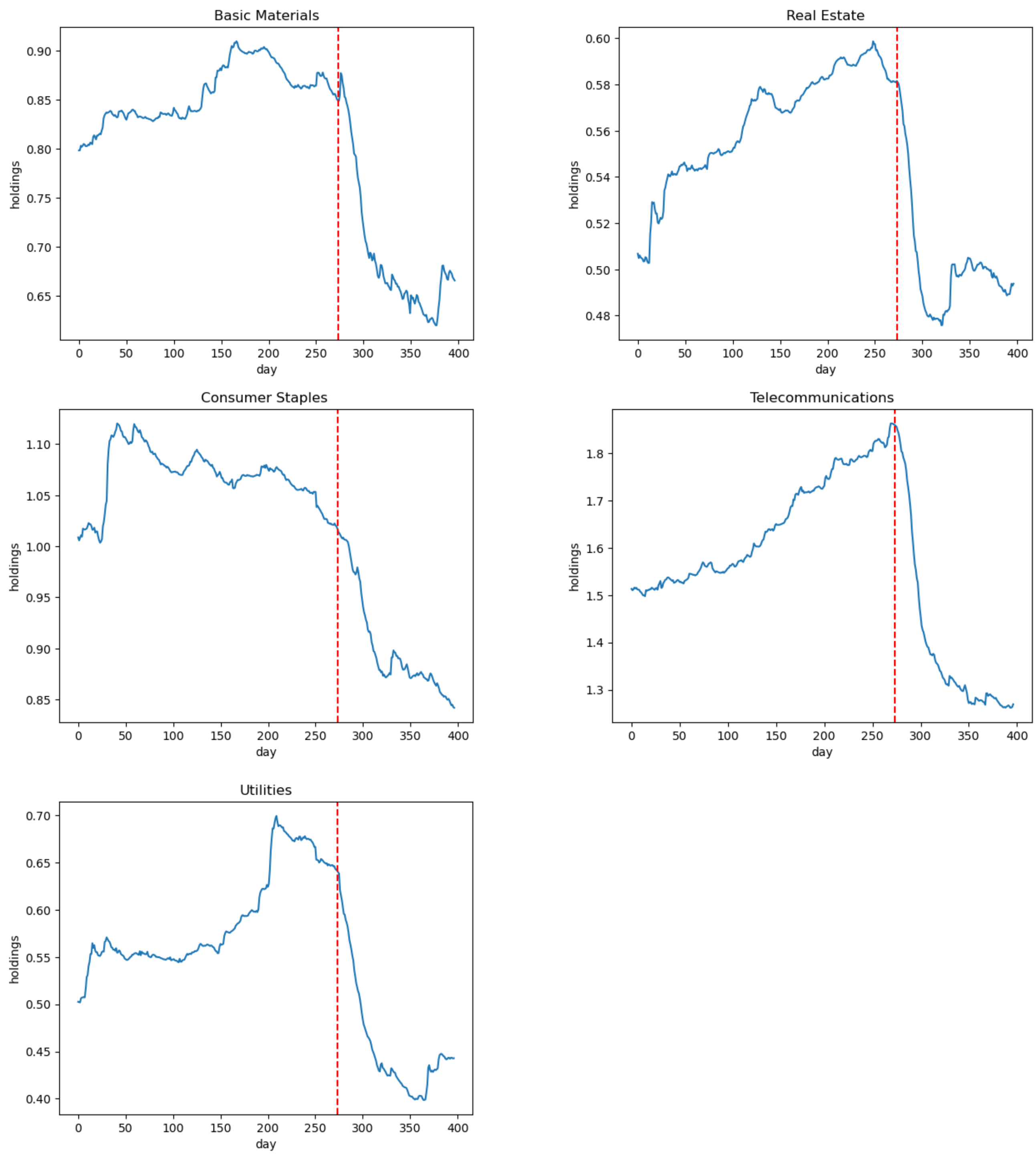


Figure 8 (cont.)



The graphs are from the 2nd of January 2019 to the 13th of August 2020. On the x-axis are the business days. The y axis is the ratio of aggregate holdings to the number of aggregate holdings in an equal-weighted portfolio. A value of 1 here would mean that the industry has the same percentage of aggregate holdings as percentage of stocks. The red line is the day of the Dow Jones drop, the 20th of February 2020.

Figure 9: Aggregate holdings per industry, relative to a value weighted portfolio

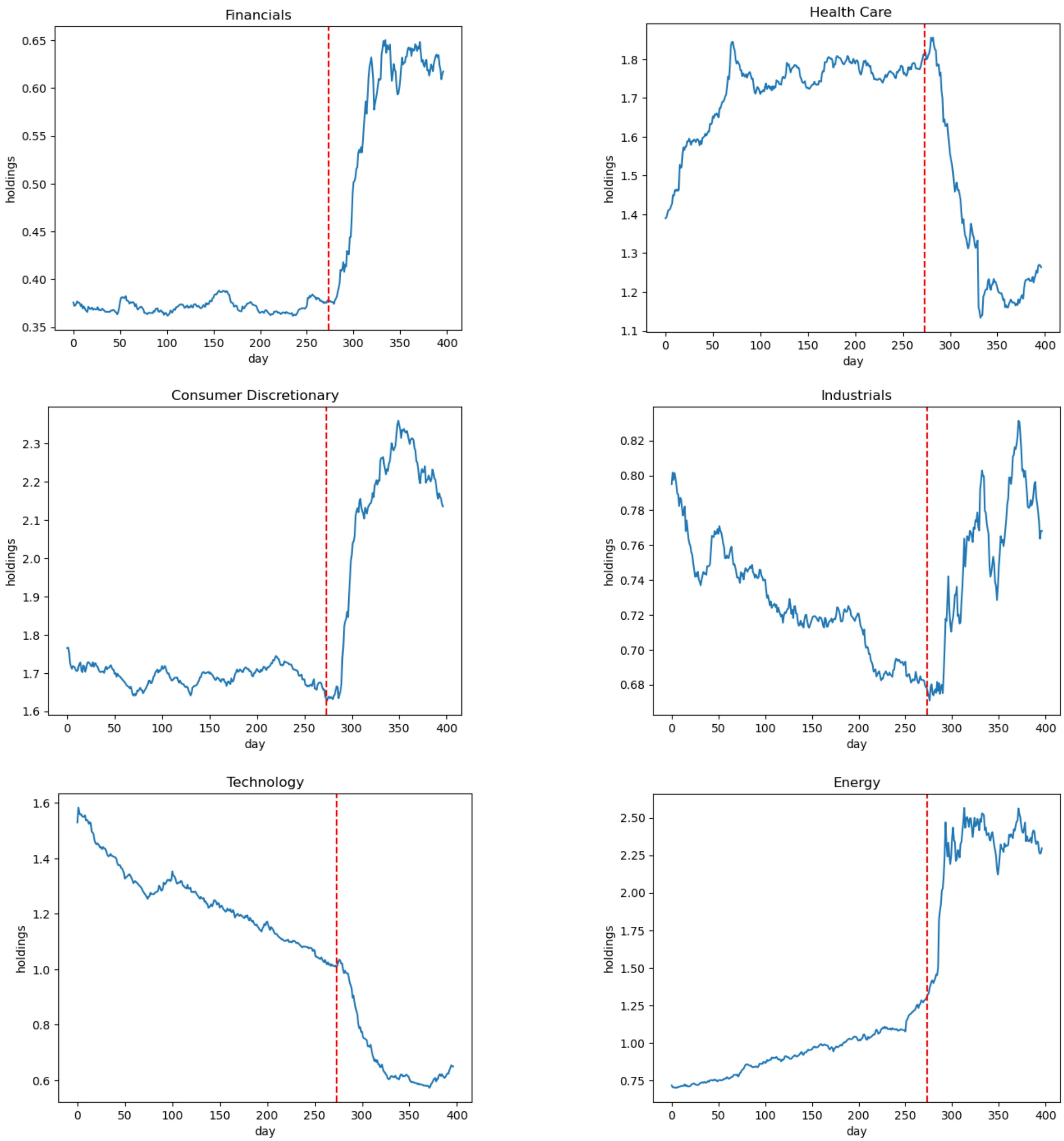
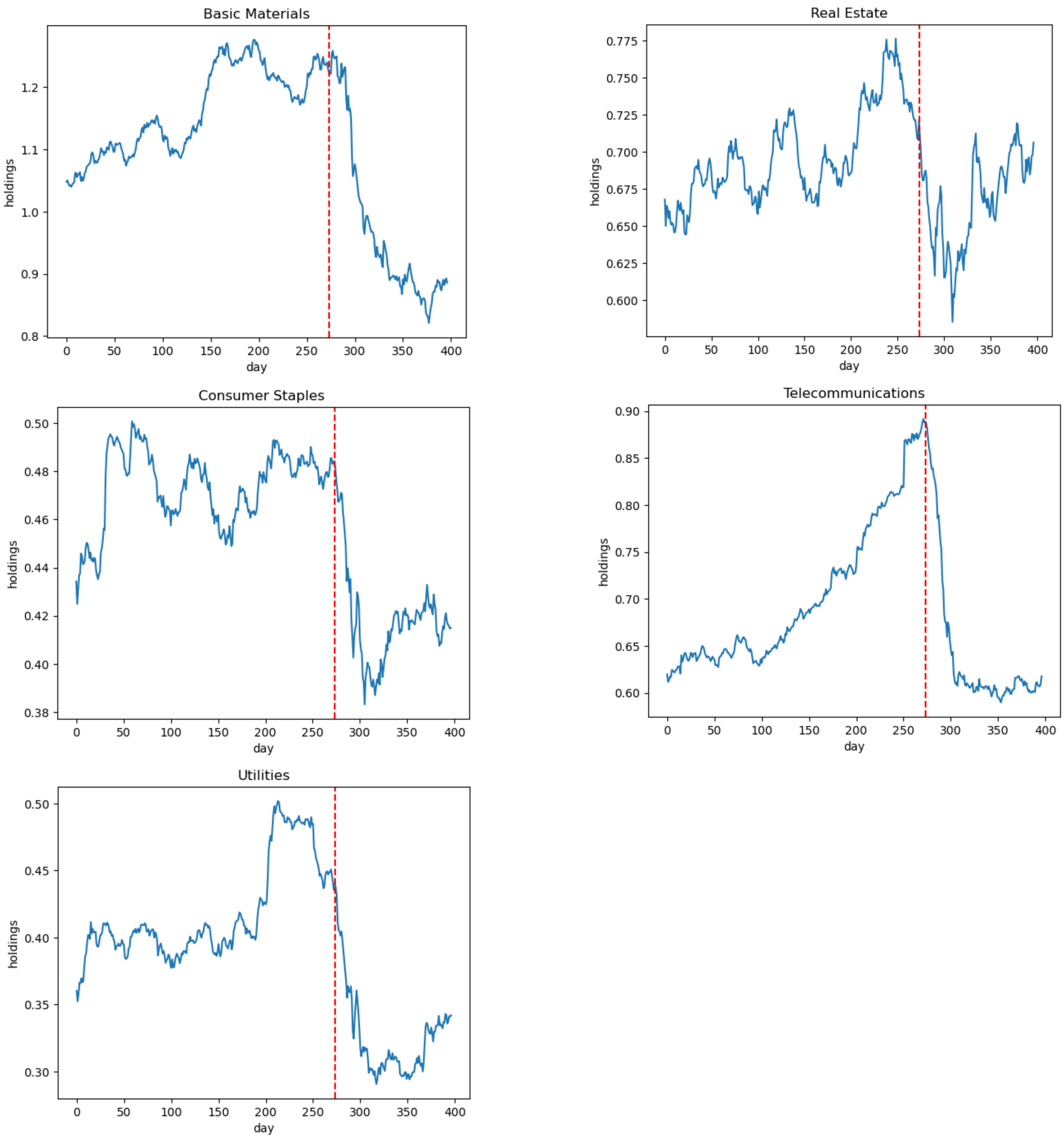


Figure 9 (cont.)



The graphs are from the 2nd of January 2019 to the 13th of August 2020. On the x-axis are the business days. The y axis is the ratio of aggregate holdings to the number of aggregate holdings in a value-weighted portfolio. A value of 1 here would mean that the industry has the same percentage of aggregate holdings as percentage of market value. The red line is the day of the Dow Jones drop, the 20th of February 2020.

Figure 10: Market value per industry, as a percentage of aggregate market value

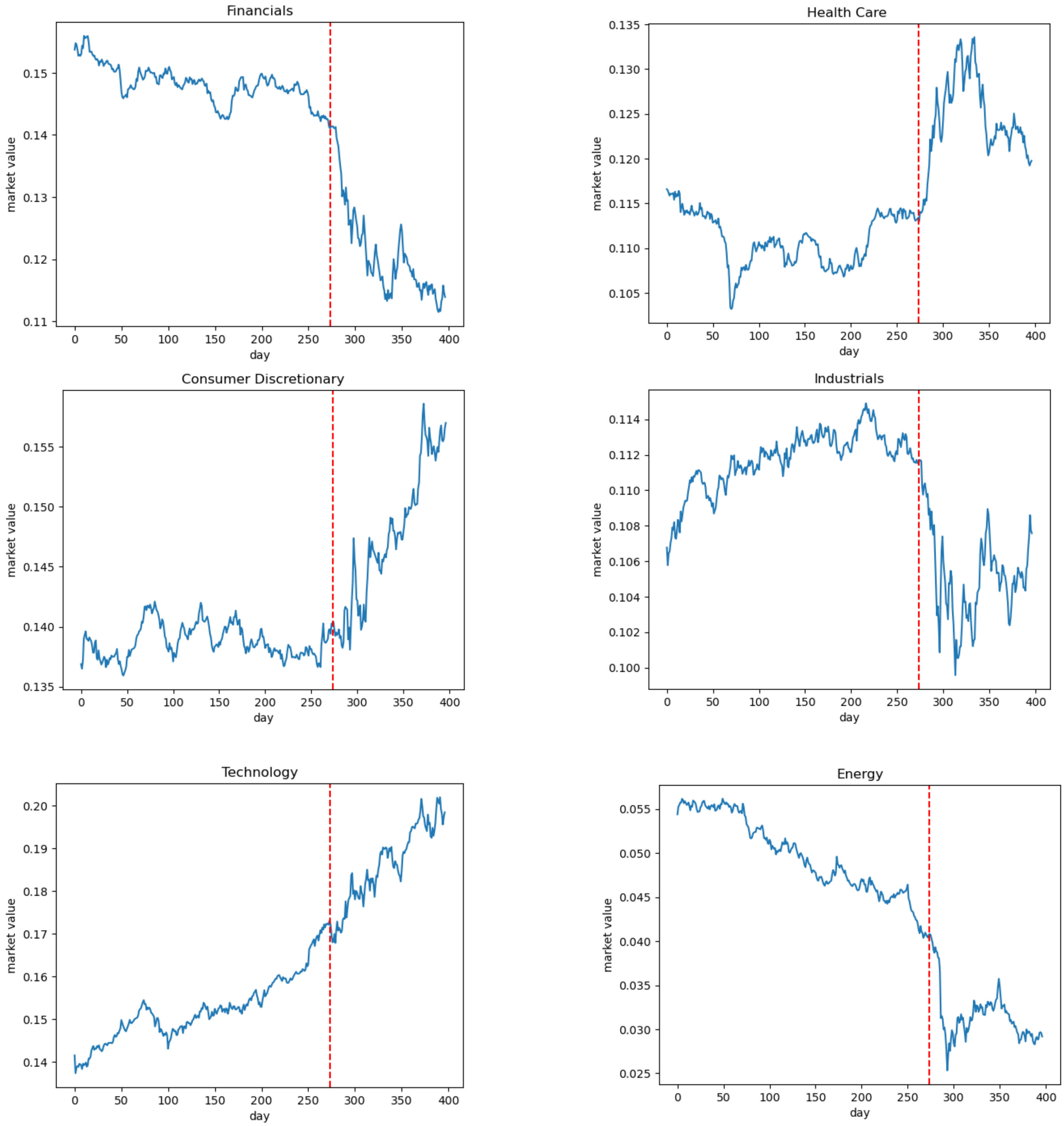
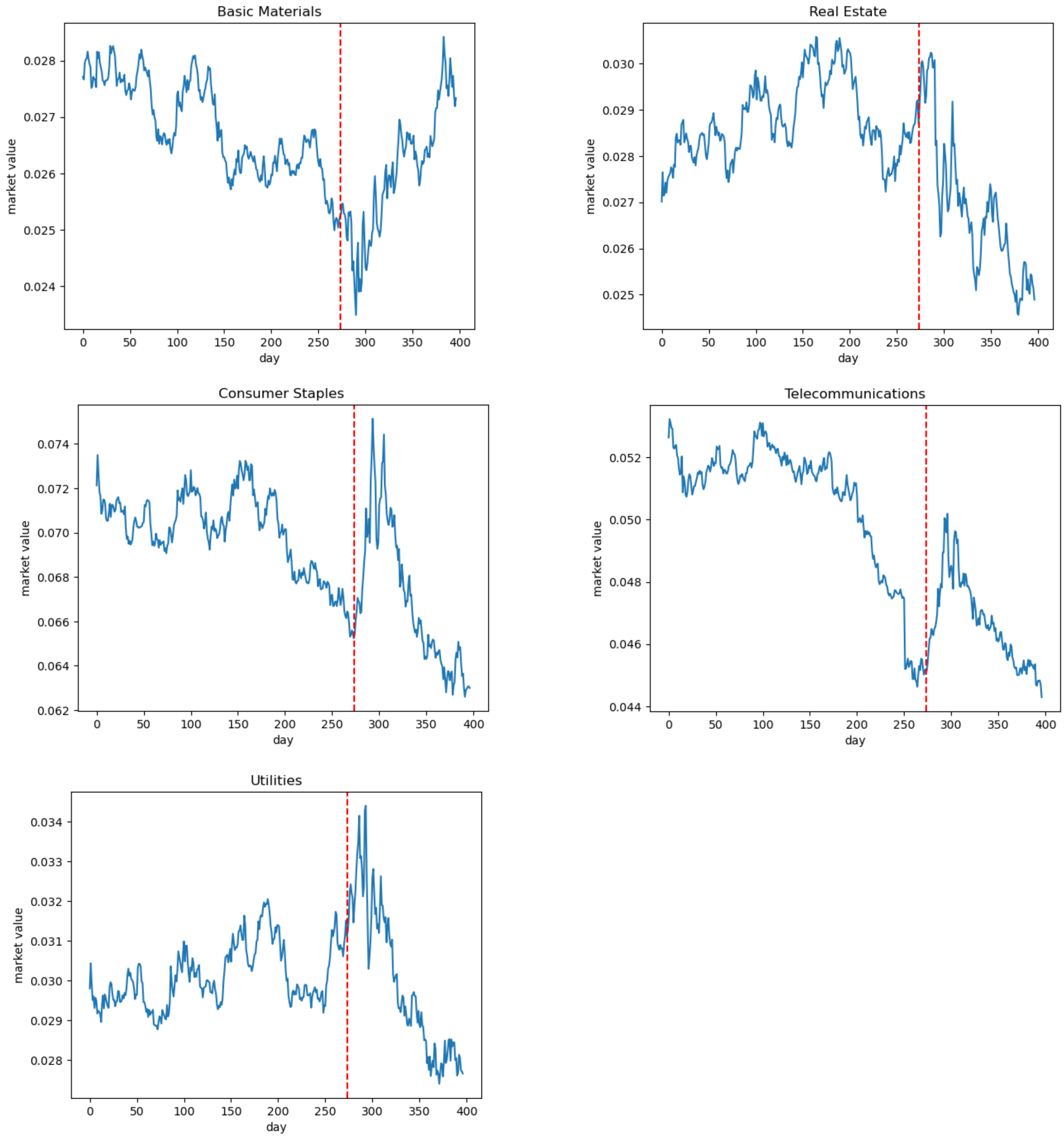
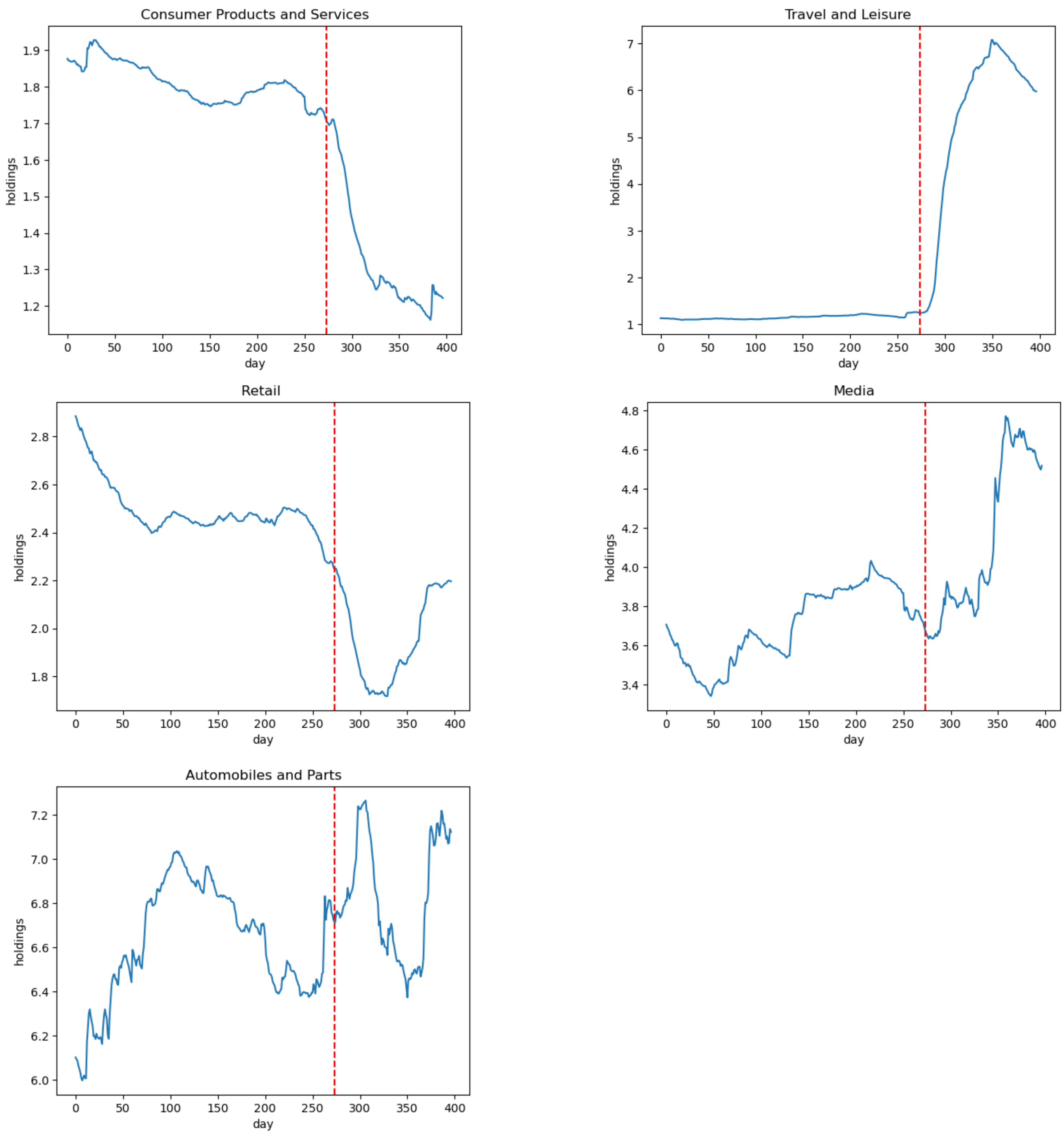


Figure 10 (cont.):



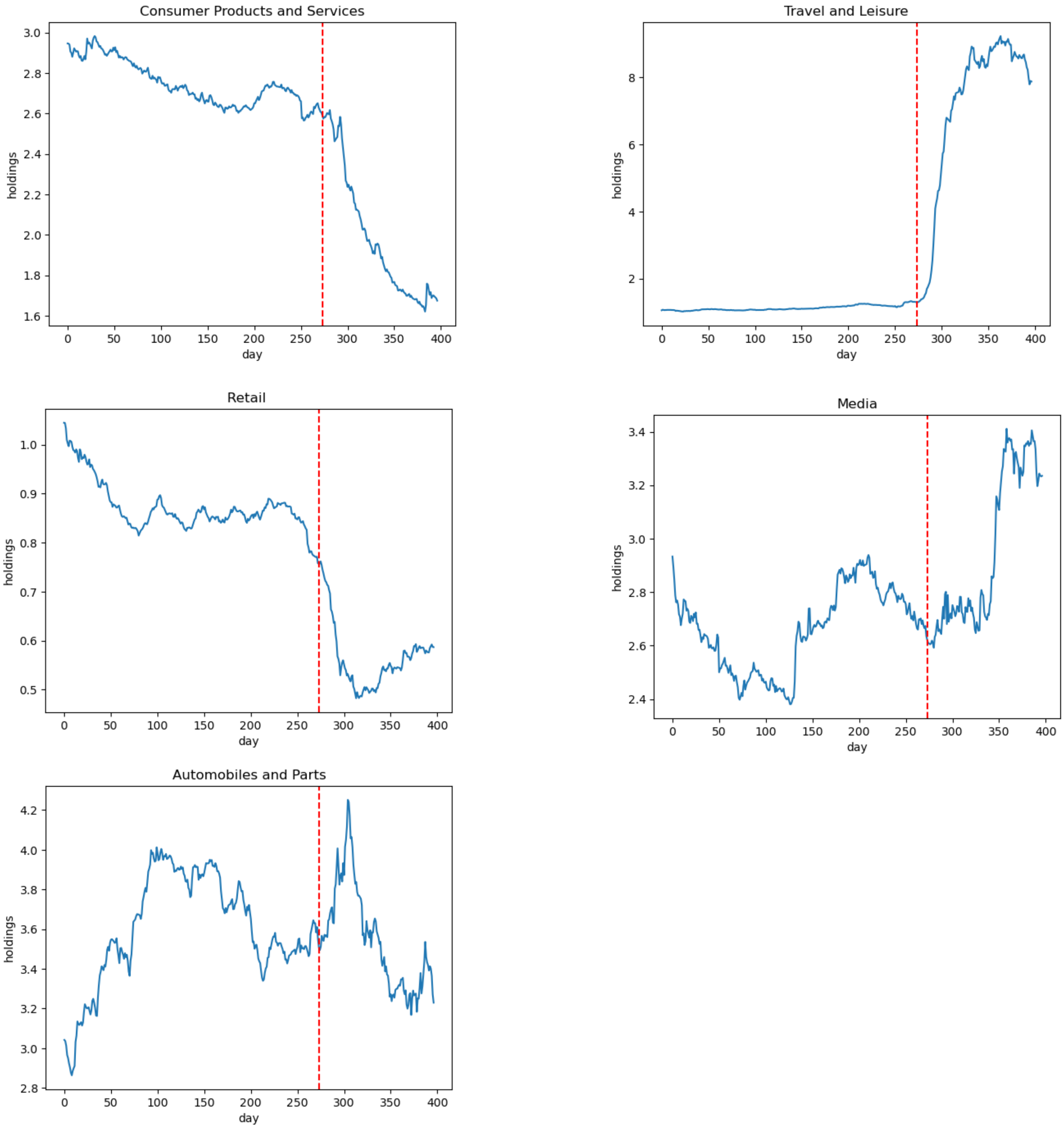
The graphs are from the 2nd of January 2019 to the 13th of August 2020. On the x-axis are the business days. The y axis is the ratio of the industry's sample market value to the sum of the market value of the entire sample. The red line is the day of the Dow Jones drop, the 20th of February 2020.

Figure 11: Aggregate holdings per Consumer Discretionary sector, relative to an equal weighted portfolio



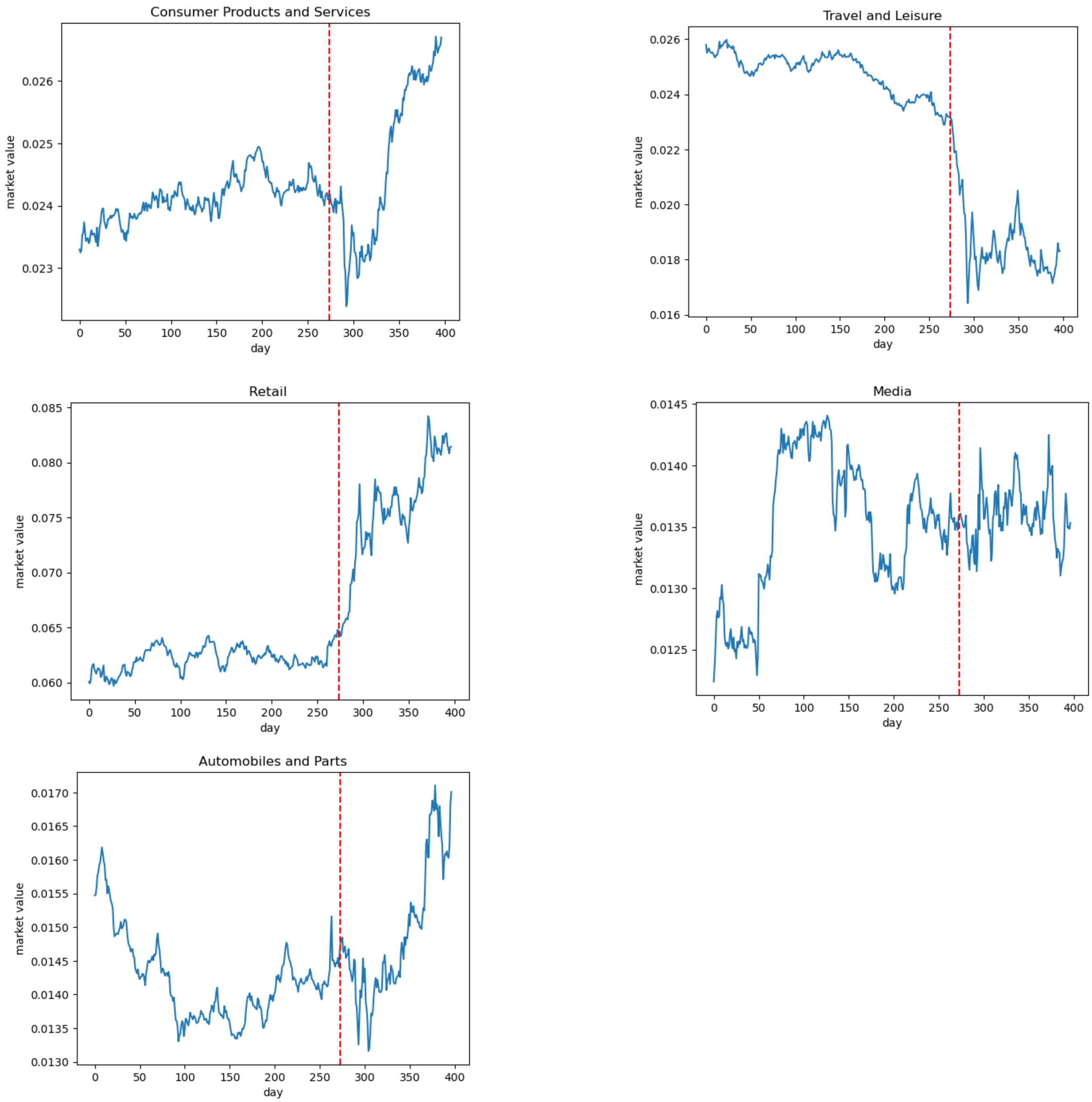
The graphs are from the 2nd of January 2019 to the 13th of August 2020. On the x-axis are the business days. The y axis is the ratio of aggregate holdings to the number of aggregate holdings in an equal-weighted portfolio. A value of 1 here would mean that the sector has the same percentage of aggregate holdings as percentage of stocks. The red line is the day of the Dow Jones drop, the 20th of February 2020.

Figure 12: Aggregate holdings per Consumer Discretionary sector, relative to a value weighted portfolio



The graphs are from the 2nd of January 2019 to the 13th of August 2020. On the x-axis are the business days. The y axis is the ratio of aggregate holdings to the number of aggregate holdings in a value-weighted portfolio. A value of 1 here would mean that the sector has the same percentage of aggregate holdings as percentage of market value. The red line is the day of the Dow Jones drop, the 20th of February 2020.

Figure 13: Market value per Consumer Discretionary sector, adjusted for the number of stocks in each



The graphs are from the 2nd of January 2019 to the 13th of August 2020. On the x-axis are the business days. The y axis is the ratio of the sector's sample market value to the sum of the market value of the entire sample. The red line is the day of the Dow Jones drop, the 20th of February 2020.