INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Does the number of Robinhood's subscribers influence the stock prices?

Pedro Miguel Machado Fernandes

Master in Finance

Supervisor:
PhD José Joaquim Dias Curto, Associate Professor with Habilitation,
Iscte-Iul

October 2021

Department of Finance

Does the number of Robinhood's subscribers influence the stock prices?

Pedro Miguel Machado Fernandes

Master in Finance

Supervisor:
PhD José Joaquim Dias Curto, Associate Professor with Habilitation,
Iscte-Iul

October 2021

Acknowledgements

I would like to thank my supervisor Professor José Dias Curto, for his support, advises and availability. This work could not have been completed without his assistance.

I also want to thank my parents, Américo e Elsa, for always motivating me, believe in me and for all the efforts taken during my academic life in Lisbon.

Finally, I would like to thank to Joana for the constant support and love.

Does the number of Robinhood's subscribers influence the stock prices?

Does the number of Robinhood’s subscribers influence the stock prices?

Resumo

O presente trabalho visa estudar a influência do aumento do número de subscritores da Robinhood, que são investidores de retalho, no preço das ações cotadas nesta corretora entre 2018 e 2020. Além disso, como a meio deste período ocorreu a pandemia da COVID-19, investigamos também a influência que o aumento do número subscritores teve nas ações exclusivamente no período após o primeiro dia de confinamento nos EUA, visto a literatura ter verificado um número crescente de investidores de retalho neste período.

Para isso, utilizámos dados em painel, com observações diárias do número de subscritores e do preço das ações. As ações que escolhemos para este estudo pertencem, essencialmente, às 20 empresas do S&P 500 com maior capitalização de mercado.

Verificámos que o aumento do número de subscritores não foi suficiente para influenciar significativamente o preço das ações da Robinhood, quer em todo o período estudado, quer a partir do confinamento nos EUA. Neste sentido, apresentamos posteriormente algumas justificações para os resultados.

Palavras-chave: Robinhood, Preço das ações, Investidores de retalho, Dados em painel.

Classificação JEL: C33 - Panel Data Models; G24 – Brokerage.

Does the number of Robinhood’s subscribers influence the stock prices?

Does the number of Robinhood’s subscribers influence the stock prices?

Abstract

The present work aims to study the influence of the increase in the number of Robinhood’s subscribers, who are retail investors, on the prices of stocks between 2018 and 2020. Furthermore, and because in the middle of this period the pandemic of COVID-19 has occurred, we also investigate the influence that the increase in the number of subscribers had on stock exclusively in the period after the first day of lockdown in the US, as the literature has verified a growing number of retail investors in this period.

For this, we used a panel data approach, with daily data on the number of subscribers and stock prices. The stocks we chose for this study belong, essentially, to the top 20 S&P 500’s companies with the largest market capitalization.

We found that the increase in the number of subscribers was not enough to significantly influence Robinhood’s stock prices, neither throughout all the studied period nor just after the first day of lockdown. In this sense, we later present some justifications for the results.

Keywords: Robinhood, Stock prices, Retail investors, Panel data.

JEL Classification: C33 - Panel Data Models; G24 – Brokerage.

Does the number of Robinhood’s subscribers influence the stock prices?

Does the number of Robinhood’s subscribers influence the stock prices?

Index

List of Tables

List of Graphs

Does the number of Robinhood's subscribers influence the stock prices?

Does the number of Robinhood's subscribers influence the stock prices?

List of Abbreviations

ADF – Augmented Dickey–Fuller

BLUE – Best Linear Unbiased Estimator

COVID-19 – Coronavirus Disease 2019

DJIA – Dow Jones Industrial Average

ETF – Exchange Traded Fund

GLS – Generalized Least Squares

IPS - Im Pesaran and Shin

LSDV – Least Squares Dummy Variable

MW – Maddala and Wu

NASDAQ – National Association of Securities Dealers Automated Quotations

NYSE – New York Stock Exchange

OLS – Ordinary Least Squares

S&P 500 – Standard & Poor's 500

SEC EDGAR – Securities and Exchange Commission Electronic Data Gathering, Analysis, and Retrieval system

US – United States

Does the number of Robinhood's subscribers influence the stock prices?

1. Introduction

Trading goods has been a part of our lives since the early civilizations. Transactions between individuals and cultures have developed to what we now call economy. In more recent times, the first modern stock trading was born in Amsterdam, in 1611. The Dutch East India Company had the genius idea to sell stocks, in order to raise capital and pay dividends to the investors. It wasn't until the 18th century that the US had a "sort of" stock market. Despite some trouble, this system has evolved and thrived. Nowadays, institutional and retail investors can not only trade stocks but also many other products.

Robinhood is an American Company that was launched in 2013 and provides financial services through a website and a mobile app. It was the first to offer zero-fee trading, offering the possibility to invest in stocks, ETFs and other financial products without any commission, thus gaining a lot of popularity. The platform, which is intuitive to use, has some helpful functionalities, for instance, it creates a personalized portfolio for each user with its companies and funds; it sends customized notifications and it presents lists for guidance, such as "100 Most Popular" which shows the 100 stocks with the highest number of Robinhood subscribers.

It has always been questioned whether these retail investors play an influential role in the market or not. On one hand, each one of these retail investors invests a small amount of money when compared with big investors, moreover, they don't have the expertise that the big investors have; but on the other hand, there are increasingly more people doing small investments, and a large number of small investments may lead to a change big enough to be noticed in the market. By analyzing the graphics presented in www.robintrack.net (Annex A), which is a website that has provided information about the trading being made in Robinhood, it has come to our attention that there might be a relation between the number of Robinhood subscribers and stock prices.

This question is especially interesting in lockdown, due to government restrictions to prevent the spread of COVID-19, a life event that has never happened before. With much of the population of the US inside their homes for a long period of time, with no outside entertainment happening and with more free time than ever, apps like Robinhood can be appealing. Either for entertainment or for money making or both. Actually, there was a significant increase in Robinhood users since the lockdown, as can be seen in the graphics (Annex A). So, the growing number of users is the best environment to study the relation between Robinhood subscribers and the stock prices.

In order to study this influence, we used a panel data approach, sometimes referred to as longitudinal data. Panel data combines both the observations collected chronologically at a regular frequency like time series data and observations across a collection of individuals like cross-sectional data. Repeated observations on the same unit allows economists to draw more sophisticated and realistic models.

The present study is divided into chapters that break the work into distinctive sections. Firstly, we will present the literature review in which we explore articles whose themes resemble that of this present study and help contextualize it. We write about the COVID-19 phenomenon and how it shaped the financial markets as well as about the spike in retail trading, their influence and their role on liquidity. Following, there is the chapter of the methodology in which we write about the approach we have chosen to answer our main question, which is, as mentioned before, a panel data approach. We talk about what it is and its advantages. Moreover, we explore the models that could be applied to our research and all the necessary tests to achieve a reliable conclusion. Afterwards, we present our hypothesis and our empirical model. Our first plan was to investigate the general impact of the number of Robinhood's subscribers on stock prices, however looking at the preliminary results incentivized us to study the same question but under two distinctive time restrictions: one with data from before the COVID-19 lockdown and one with data after it. We decided to divide the data this way in order to investigate exactly how Robinhood's subscribers impact the stock prices, that is, if the effect is verified only before or only after the lockdown, or, otherwise, if it is verifiable regardless the lockdown. The next chapter presents our data and empirical results. In this chapter, we present the 20 companies chosen for this study as well as the statistical procedures in R made in order to verify our hypothesis. We then discuss our results under the chapter "Conclusion" as well as present the limitations of the present study and, consequently, suggestions for future research.

Does the number of Robinhood's subscribers influence the stock prices?

2. Literature Review

2.1. Robinhood

Robinhood is a brokerage firm that provides its services through a trading app that was launched in 2015. It was the first platform to offer zero-fee trading, which caused its popularity. Both the app and the webpage are simple to use. Users can see their current account standing right at the home page, they can also search for firms and see graphics of historic prices and other relevant information. Not only that, but Robinhood also displays lists such as "Top Movers" which presents the 10 securities with highest and lowest returns and "100 Most Popular" which presents the 100 stocks with the highest number of investors. New entries on this last list will receive a great amount of attention from investors, and therefore are far more likely to be purchased the days following its entry in the list (Stein, 2020). According to Stein (2020), recent stocks on this list are between 3.5 and 7 times more likely to be purchased by Robinhood's investors than other stocks available at the platform.

As mentioned in Robinhood's website (Robinhood, 2021), the platform's goal is to "make investing in financial markets more affordable, more intuitive, and more fun, no matter how much experience you have". It is still commission-free and continues to grow in popularity.

2.2. Influence of retail investors in the market

Retail investors are non-institutional investors, that is, every person who invests with its own money, not on someone else's behalf. These retail investors are usually driven by personal goals such as savings. Robinhood is one of the many brokers that target this type of investors.

Despite investing a short amount of money when compared to big investors, these retail investors can make an impact in the market. Actually, trading by not fully rational traders can drive prices away (Barber, Odean & Zhu, 2008). When non-informed traders buy, assets do become overpriced and when they sell, assets become underpriced. Nonetheless, asset prices eventually revert to their fundamental values (Barber, et al., 2008). Barber et al. (2008) concluded that retail trade imbalances forecast future returns, both in short and long term. They also concluded that over short term (days) retail traders move stock prices in the direction of their trades, but in the long term (approximately a year) they can only achieve this in small stocks. This means that the little investors do shake the market a little bit and are likely to be influential, particularly in small stocks. Abudy (2020) examined the possible costs that retail trading could have on the market quality and found that retail investment does not induce price noise in stocks. The group of retail investors is not uniform and exposing it to more financial knowledge could actually benefit the market.

2.3.What influences retail investors?

Merton (1987) was the first to introduce the concept of investor recognition and suggested that investor's attention may be an important key to stock prices and liquidity. More recently, Barber and Odean (2008) noted that stocks that were in news headlines, that were abnormally traded or that had extreme returns, caught attention of retail investors. Stocks that acquire high levels of attention do become more traded than stocks receiving less attention.

However, there is a problem with investors' attention, and that it is limited. Because of the limits of information-processing capacity of investors, as they are humans, their attention is largely on stocks that both interest them and sound familiar (Ding & Hou, 2015). Actually, the literature has shown that retail investors tend to invest in favor of firms that they are familiar with, being undoubtedly biased (Cao, Han, Hirshleifer & Zhang, 2011). Individual investors don't usually obtain access to professional information about firms, that could be found, for example in Reuters and Bloomberg, but rather to common information displayed on the internet. Ding and Hou (2015) searched the frequency data on S&P 500 stocks provided by Google Trends, as a direct measure of investor attention, between January 2004 and December 2009, and examined its impact on the shareholder base and stock liquidity. The authors concluded that this attention significantly enlarges the shareholder base and improves stock liquidity. Aouadi, Arouri & Teulon (2013) verified not only that investors' attention, measured by Google search volume, drives the stock market liquidity significantly but that investors' attention is a significant determinant of the stock market volatility.

Liaukonytė and Žaldokas (2019) studied the impact that TV advertising had on financial information acquisition, through a database of the SEC EDGAR. They found that TV advertisement cause an increase in searches for financial information of advertised companies and their rivals, which lead to a raise at the trading volume level on the advertiser's stock, mostly induced by retail investors. On the other hand, episodes of sensational news, that are exogenous to the financial markets, distract retail traders. On these "distraction days" trading activity, volatility and liquidity decreases (Peress & Schmidt, 2020).

Does the number of Robinhood's subscribers influence the stock prices?

Since retail investors do have some power on the market, this means these studies show that the stocks who receive more attention from traders are more likely to be further traded. However, the media and the internet are not the only ones that influence retail investors' decisions. Brokers also play a big role in that. According to Barber, Huang, Odean and Schwarz (2020), Robinhood users are more likely to be influenced when buying stocks because most of them are first time investors without a clear strategy. This combined with the fact that Robinhood's app supplies limited information and eases the process, makes investors rely more on their intuition and make the decision to invest in a much easier way. Thus, Robinhood's users have a greater tendency to herd, compared with other investors, which itself is influenced by the information of Robinhood's app. These factors lead to a trading concentration of Robinhood's users which impacts pricing. This happened because the number of Robinhood's users has been increasing, getting to 13 million in May of 2020. Moreover, they are more active and, as stated before, play a role on the influence on the purchases of the other retail investors, having a positive correlation with them (Barber et al., 2020).

Stein (2020) has done a similar study to ours about the 5 biggest factors that contribute to stocks entering the "100 Most Popular" list on Robinhood. This list appears on Robinhood's app, and, as the name indicates, lists the 100 stocks with the most subscribers on the app. Stein (2020) noticed that when stocks entered this list, they received a lot of investors' attention which induced a rise of subscribers of those stocks and, consequently, a rise in their prices. However, this effect has shown to be short lived, and as soon as the novelty effect disappeared, stock prices decreased again.

2.4. Retail investors and liquidity

Besides finding that Google search volume is a reliable proxy of investors' attention, Aouadi et al. (2013) also found that investors' attention is strongly correlated with trading volume. Moreover, it drives significantly the stock market liquidity and is a significant determinant of volatility. This influence was still relevant after controlling for a potential financial crisis effect. Abudy (2020) found that retail trader's participation in trading contributes to aggregate stock market liquidity. Regarding liquidity at the stock level, Abudy (2020) found that when retail traders act as buyers, they require a higher level of liquidity to trade when compared to their trading activity as sellers. Exploration of the determinants of retail investors revealed that when retail investors with higher trading frequencies appear in the market, they don't require a high level of stock market liquidity. On the other hand, when more diversified retail investors appear in the stock market, they require a higher level of liquidity in order to trade, on average.

2.5.The COVID-19 pandemic crash of the stock market

It was in December of 2019 that the contagious Coronavirus disease (COVID-19) was first identified. Due to its easy transmissibility, it quickly spread from China to the rest of the world, which led country's governors to force lockdowns. These lockdowns mostly consisted of confining the majority of the population at home. Most people were working from home if possible or not working at all and were only leaving the house to buy essentials needs or for urgent medical treatments. Most countries, including the US, entered the first lockdown in March of 2020.

This phenomenon has drastically impacted the financial markets. Moreover, it grew so quickly that almost no precautions could be made to save the market, as this pandemic was a global exogenous shock that did not come from changes in economic conditions, making it unlikely for institutional investors to predict it and avoid the stocks that would be hit the hardest (Glossner, Matos, Ramelli & Wagner, 2020). The collapse of stock prices in March was, indeed, one of the biggest stock market crashes in history. According to Dow Jones Industrial Average (DJIA) the market fell 26% in four days (Mazur, Dang & Vega, 2020). Mazur et al. (2020) showed that firms operating in the crude petroleum, real estate, entertainment and hospitality sectors lost considerably more than 70% of their market capitalization. The authors show that these stocks have more asymmetric movements and exhibit extreme volatility, which negatively correlates with stocks returns. On the other hand, natural gas, food, healthcare and software sectors performed well and generated high returns during this market crash. This crash was mirrored with a sharp increase in volatility.

On the other hand, with this new reality of staying at home, Americans felt more attracted to low-cost trading services like Robinhood (Pagano, Sedunov and Velthuis, 2020). Since January of 2020 and with a higher rate since the beginning of March, Robinhood's retail trading accounts have displayed a growing time trend (Ozik, Sadka and Shen, 2020). People that became Robinhood's subscribers and that started to invest in stocks for the first time, make up more than 50% of their customers, with the median customer age of 31 years old, according to Robinhood. This became evident when, in March of 2020, with the aggressive fall of stock price, Robinhood observed an increase of three times more on its average customer trading volume than in 2019 (Rooney, 2020). Moreover, people started to pay attention to statistics related to COVID-19 and to the stock market. From the middle of February to March of 2020, the rate per stock of an average COVID-19 media coverage increased from 29% to 72%. So, the already great availability of fintech trading platforms plus the growth of savings, lead to an increase in the retail stock market activity (Ozik et al., 2020).

During the COVID-19 market crash, US stocks with higher institutional ownership had a "higher fall", collapsing more than other stocks. As stated by Glossner et al. (2020), "Overall, the results suggest that when a tail risk realizes, institutional investors amplify price crashes by fire-selling and seeking shelter in ''hard'' measures of firm resilience.". This means that institutional investors, which play a crucial role in stock markets, amplified the market crash by choosing a risk-averse investment strategy, shifting their portfolios to firms with high cash holdings and low leverage. According to the Investment Company Institute, investors have taken out more than $150 billion from the US domestic equity funds since the beginning of 2020. Ozik et al. (2020) suggest that, despite the injection programs of Federal Reserve, the flattening of the illiquidity curve is mostly due the increase in retail trading activity during the lockdown period, what provoked the lowering stock bid-ask spreads.

In contrast, retail investors took the opposite strategy, being more interested in stocks with low cash holdings and high leverage (Glossner et al., 2020). These preferred extreme recent winners and losers in a consistent way, indicating an absence of panic and margin calls (Welch, 2020). This behavior continues to be observed in the second quarter of the year (Glossner et al., 2020). Some major brokerage firms saw the number of new accounts being created increased to a level never seen before, with retail trading spiking up due to the lock-down (Ozik et al., 2020). Glossner et al. (2020) claimed that with this increase, retail investors acted as liquidity providers.

This effect, as expected, was also noticed in Robinhood, as we can observe in Welch's (2020) working paper: "… there is evidence that as the stock market declined, investors actively added cash to fund purchases of more stocks. … Thus, the evidence suggests that RH investors collectively acted as a (small) market-stabilizing force." (Welch, 2020).

Based on these studies, and according to our main objective, the purpose of this research is to investigate whether this recent rise in Robinhood's subscribers leads to a significant rise in stock returns.

Does the number of Robinhood’s subscribers influence the stock prices?

Does the number of Robinhood's subscribers influence the stock prices?

3. Methodology

In this chapter, we will firstly explain the theoretical models that are going to be used in this work and then present both our hypothesis and the multiple linear regression.

3.1. Panel data

With the advance of technology, it has been possible to store and work with huge stacks of data. Thus, models based on panel data are of great use. A panel data collects, over a specific time frame, repeated observations of the same units, as in the present paperwork, stock prices and number of Robinhood's subscribers of those stocks. This allows economists to estimate more complex and realistic models rather than working with a single cross-section or single time series, as panel data usually refer to a combination of at least two of these dimensions (Verbeek, 2012).

In panel data, the static linear regression model can be written as

$$y_{it} = \alpha_i + \lambda_t + x'_{it}\beta' + \varepsilon_{it}, \tag{1}$$

where α_i represents the individual-specific effects, which means it corresponds to the effects of variables that are specific to the i^{th} individual and are constant over time; λ_t represents the time-specific effects; β is a $K \times 1$ vector of parameters; and x_{it} represents a K-dimensional matrix of explanatory variables. Since we are talking about panel data, the variables are indexed with an *i* for the individual (i = 1, …, N) and a *t* for the time period (t = 1, …, T) (Hsiao, 2014).

3.2. Advantages of Panel Data

As stated before, panel data allows for a time-series observation of several units, and therefore leads to more realistic models. These realistic models reveal some advantages when compared to cross-section and time-series models. For instance, panel data is best suited for the analysis of changes over time, namely, at the individual level. This model is adequate for generating more accurate predictions for individual outcomes. Other advantage is that it gives a large number of data points, thus allowing for more degrees of freedom, as more sample variability. This ultimately reduces collinearity between explanatory variables which leads to a more efficient econometric estimates (Baltagi, 2005; Hsiao, 2007).

As panel data allows the study of more complex economic issues, it is suited to the analysis of hypotheses with higher complexity. The mix of interindividual and intraindividual differences allow the analysis of a number of important economic questions that otherwise, with cross-sectional or time series data sets, could not be answered. Thus, panel data is suited for the study of individuality throughout time. That is, is not only suitable to model or explain why individual units behave differently but also to model why a given unit behaves differently at different time periods (for example, due to a different past) (Hsiao, 2014). Another advantage is that the model better controls the impact of omitted variables that correlate with the explanatory variables. This control is due to the fact that panel data utilizes information on both the intertemporal dynamics and the individuality of the entities investigated. To conclude, panel data is adequate to uncover microdynamic and macrodynamic relationships and effects that couldn't be estimated using another model (Hsiao, 2014).

3.3. Static Linear Model

Now, we will be discussing the static linear model of a panel data setting. Firstly, we will present the fixed effects model and the random-effects model, and then we will discuss which one of the models is best suited for our research.

Finally, we would like to mention that, in our panel, the individual unit is observed at all time frames, leaving no space for non-observed values and, thus, making it a balanced panel.

3.4. The Fixed Effect Model

The fixed-effect model is a simple linear regression model, specified as bellow:

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{ti}, \qquad \varepsilon \sim IID(0, \sigma_\varepsilon^2) \tag{2}$$

where α_i is a N x 1 vector; β is a 1 x K vector of constants and ε_{ti} is the error, thus representing the omitted variables. It is often assumed that all x_{it} are independent of all ε_{ti} (Verbeek, 2004).

In this model, the intercept terms vary over the individual units, thus allowing the individual effects (that can't be observed) to correlate with the chosen variables. This model doesn't present significant temporal effects, but it does present significant differences in the cross-sectional level. The dependent variable y_{it} depends on K exogenous variables - x_{it}- which differ among individuals at a cross-section, at a specific point in time, and demonstrate variations throughout time (Verbeek, 2004).

This model can also be written in the usual regression framework, which includes dummy variables - one for each unit *i* in the equation - as shown below. These variables allow for individual effects (Verbeek, 2004).

Does the number of Robinhood's subscribers influence the stock prices?

$$y_{it} = \sum_{j=1}^{N} \alpha_j d_{ij} \; + \; x'_{it}\beta \; + \; \varepsilon_{it} \tag{3}$$

where d_{ij}= 1 when i = j, otherwise 0. This way, we have a set of N dummy variables in the model. In this equation, the parameters $\alpha_1, \ldots, \alpha_N$ and β are estimated by ordinary least squares (OLS). This last parameter - β- is called the "least squares dummy variable" (LSDV) estimator. As this regression model has quit a few regressor, it is useful and possible to compute the estimator for β in a simpler way. This estimator remains the same when the regression is performed in deviations from individual means, which means that the individual effects - α_i- are eliminated (Verbeek, 2004). Firstly, we must transform the data, as seen below

$$\bar{y}_i = \alpha_i + \bar{x}'_i\beta \; + \; \bar{\varepsilon}_i, \tag{4}$$

where $\bar{y}_i = T^{-1}\sum_t y_{it}$, and consequently for the other variables. Then, it can be written as

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)'\beta \; + \; (\varepsilon_{it} - \bar{\varepsilon}_i) \tag{5}$$

Here, the transformation that produces observations in deviation from individual means is called "within transformation". Thus, the OLS estimator for β obtained in this model is not only called the "fixed effect estimator", but also the "within estimator" (Verbeek, 2004). This estimator is identical to the LSDV estimator, and it can be given by

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)' \tag{6}$$

The fixed effect estimator is unbiased for β, when assumed that all x_{it} are independent of all ε_{it}. Moreover, $\hat{\beta}_{FE}$ has a normal distribution when the normality of $\varepsilon_{it}$is imposed. For consistency, it is required that

$$E\{(x_{it} - \bar{x}_i)\varepsilon_{it}\} = 0 \tag{7}$$

Considering that x_{it} and $\bar{x}_i$ are uncorrelated with the ε_{it}, it is implied that

$$E\{x_{it}\varepsilon_{is}\} = 0 \qquad for \; all \; s, t \tag{8}$$

which is called strictly exogenous (Verbeek, 2004).

As explanatory variables are independent of all errors, the N intercepts can be unbiasedly estimated as:

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}_{FE}, \qquad i = 1, \ldots, N. \tag{9}$$

Moreover, ε_{it} being independent and identical distributed across both individuals and time with variance σ_ε^2, the covariance matrix for the fixed effect estimator $\hat{\beta}_{FE}$ is

$$V\{\hat{\beta}_{FE}\} = \sigma_\varepsilon^2 \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \quad (10)$$

A consistent estimator for σ_ε^2 is the within residual sum of squares divided by N(T-1), as we can see here:

$$\hat{\sigma}_\varepsilon^2 = (N(T-1))^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_i - (x_{it} - \bar{x}_i)'\hat{\beta}_{FE})^2 \quad (11)$$

The fixed effect estimator could be asymptotically normal under weak regularity conditions, and so the usual inference procedures, such as t and Wald tests, can be used.

To conclude, the fixed effects model explains to what extent y_{it} differs from $\bar{y}_i$, not focusing on why $\bar{y}_i$ is different from $\bar{y}_i$. This means that this model is based on differences "within" individuals. It must not be forgotten that the parametric assumption made about β impose that a change in x could derive from a change in period or in individual, both having the same effect. It must be also noted that, because the fixed effects estimators only depend on deviations from their group means, the results from a fixed-effect regression have to be interpreted with a mindset aware of the fact that the parameters are identified only through the within dimension of data (Verbeek, 2004).

3.5. The Random Effects Model

The random effects model is adequately used when the individual effects are strictly uncorrelated with the regressors. In regression analysis it is usually assumed that all the factors that affect the dependent variable, but have not been included as regressors, can indeed be address by a random error term (Verbeek, 2004). In this model, the α_i are random factors that are considered in the "random error" term (Verbeek, 2012; Wooldridge, 2017). The model can be written as:

$$y_{it} = \beta_0 + x_{it}'\beta + \alpha_i + \varepsilon_{it},\ \varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2);\ \alpha_i \sim IID(0, \sigma_\alpha^2). \quad (12)$$

These α_i are independently and identically distributed over all factors that affect the dependent variable. Here, α_i is constant over time and ε_{it} is assumed to be uncorrelated over time. The OLS estimator for β_0 and β is expected to be unbiased and consistent, since it is assumed that α_i and ε_{it}

are not only mutually independent but also independent of all x_{js} (Verbeek, 2004). However, the composite error term $\alpha_i + \varepsilon_{it}$ has some autocorrelation which turns the OLS estimators incorrect.

Considering $v_{it} = \alpha_i + \varepsilon_{it}$, the variance matrix of the error term is:

$$\Omega_{vi} = V(\alpha_i|X_i) = \begin{bmatrix} \sigma_v^2 & \dots & \sigma_\alpha^2 \\ \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \dots & \sigma_v^2 \end{bmatrix} \tag{13}$$

where $\sigma_v^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$. This case where the variance-covariance matrix of the error term is not diagonal is called the equi-correlated random effects model (Schmidheiny, 2012). The diagonal elements of the variance covariance matrix of the error term are:

$$E(v_{it}^2) = E(\alpha_i + \varepsilon_{it}) = E(\alpha_i^2) + E(\varepsilon_{it}^2) + 2cov(\alpha_i, \varepsilon_{it}) = \sigma_\alpha^2 + \sigma_\varepsilon^2 \tag{14}$$

Despite the least squares estimator still being consistent, it is no longer BLUE. This means we are in the presence of a non-spherical variance-covariance matrix of the error term problem. To solve this problem, the generalized least squares (GLS) estimator can be used instead of the OLS estimator, since it takes into account non-spherical disturbances (Kenkel, 2021). The GLS estimator is the combination of the fixed effects estimator and the between estimator. With this said, the GLS estimator can be written as:

$$\hat{\beta}_{GLS} = W\hat{\beta}_B + (I_K - W)\hat{\beta}_{FE} \tag{14}$$

where I_K is the K-dimensional identity matrix, W is the weighting matrix, $\hat{\beta}_{FE}$ is the fixed effects estimator and

$$\hat{\beta}_B = \left(\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\right)^{-1} \sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})' \tag{15}$$

is the between estimator for β.

3.6. Pooled OLS vs Fixed Effect vs Random Effect

In the pooled model, the data is pooled all together and individual effects average out. This regression assumes a common intercept β_0 for every cross-sectional units and exogenous regressors x_{it}, not including any fixed or random effects. The composite error is $u_{ti} = (\alpha_i - \bar{\alpha} + \varepsilon_{it})$ in which individual effects $\alpha_i - \bar{\alpha}$ are approximately 0 and the idiosyncratic error $\varepsilon_{it} \sim i.i.d.(0, \sigma^2)$ (Cameron & Trivedi, 2010). The pooled model can be written as:

$$y_{it} = \beta_0 + x'_{it}\beta + u_{ti}, \qquad \text{for i} = 1, \dots, \text{N et t} = 1, \dots, \text{T} \tag{16}$$

Fischer test (F test) is used to test which model, pooled or fixed, fits better the data. The null hypothesis of this test states that all individual intercepts are equal to zero. Rejecting this hypothesis means that the fixed effects are non-zero and therefore, the fixed effect model is the more adequate.

To choose between the pooled and random effect models, the Breusch-Pagan Lagrange Multiplier test is an adequate test to use. The null hypothesis of this test states that all individual specific variance components are zero and the rejection of this hypothesis means that there is random effect. Thus, the random effect model is preferred to the Pooled OLS model.

When T is large, both the LSDV and the GLS estimators become the same estimator. In this case the choice between a fixed effect and a random effect isn't problematic since both work. It is when T is small and N is large that the choice must be wise (Hsiao, 2014).

In order to choose from the random effects model or the fixed effects model, the Hausman test (Hausman, 1978) can be useful. This test compares two estimators: one that is consistent under the null hypothesis and under the alternative hypothesis and one that is only consistent under the null hypothesis. The null hypothesis is unlikely to hold when there is a significant difference between these two estimators. The null hypothesis states that the random effects is the preferred model because it saves N-1 degrees of freedom when compared with fixed effects and therefore secures more efficient coefficient estimates (Verbeek, 2004). Not only that, but it can estimate time-invariant regressors as well, as opposed to fixed effects (Cameron & Trivedi, 2010). On the other hand, rejecting the null leads to choosing the fixed-effects model.

3.7. Autocorrelation and Heteroskedasticity

Two of the most impactful problems that can arise with panel data models are heteroskedasticity and serial correlation (or autocorrelation). When one or both of these two are present in the error term, the statistic conclusion will be misleading. With this said, it is crucial to detect autocorrelation and heteroskedasticity using the adequate tests.

When talking about the fixed effect model, the autocorrelation and heteroscedasticity tests are quite simple to execute, since the model is mostly estimated by OLS. On the other hand, when talking about the random-effects model, the computation of these tests is more complicated (Verbeek, 2004). However, it is possible to use the tests usually applied to the fixed effect model in random effects situations when assuming that α_i is independent and identically distributed and independent of the other explanatory variables (Verbeek, 2012).

Does the number of Robinhood's subscribers influence the stock prices?

3.7.1. Autocorrelation

Autocorrelation, also known as serial correlation, occurs when there is correlation of the error term with its past values.

There are three factors that could cause autocorrelation in OLS: omitted variables, dynamic misspecification and functional form misspecification (Verbeek, 2004).

Autocorrelation is detected either by graphical representation or by statistical tests. The most common way to verify autocorrelation is with the former, in which residuals and hypothesis testing are represented. We will be using the latter in our research.

In panel data models, in order to test for autocorrelation, we can use both a modified Durbin-Watson test and a Breusch-Godfrey test. The generalization of the Durbin-Watson test is (Verbeek, 2012):

$$dw_p = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} (\hat{\varepsilon}_{it} - \hat{\varepsilon}_{i,t-1})^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2} \tag{17}$$

While the modified Durbin-Watson test is restricted to detecting first order autoregression,

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \tag{18}$$

where v_t is an error term with mean zero and constant variance σ_v^2, the Breusch-Godfrey test can detect autocorrelation to any predesignated order p, while also supporting a broader class of regressors (Verbeek, 2004). This means that this last test covers autocorrelation of higher orders whether the regressors include lags of the dependent variable or not.

The null hypothesis of both tests is:

$$H_0: \rho = 0 \tag{19}$$

The presence of autocorrelation will lead to some consequences, namely that the OLS estimators become less efficient and, therefore, no longer BLUE and that the estimated variances for the regression coefficients are biased, due to underestimation of the value of the standard errors, thus the values of t and F tests aren't valid anymore. Moreover, the value of R^2 is also overestimated, pointing to a biased better fit than it actually is. However, the estimators and forecasting from the OLS method are still valid. The inconsistency appears when there are lagged dependent variables as explanatory variables (Verbeek, 2012; Berry and Feldman, 1985).

Does the number of Robinhood's subscribers influence the stock prices?

3.7.2. Heteroskedasticity

Heteroskedasticity is when V{ε|X} is diagonal, although not equal to σ^2 times the identity matrix. This problem, usually encountered in cross-sectional models, means that the variance of ε_i may vary over the observation. This is the opposite of one of the conditions of the Gauss-Markov theorem - Error's homoskedasticity:

$$\text{var}(\varepsilon_i|X) = \text{E}(\varepsilon_i^2|X) = \sigma^2, \qquad i = 1,2,3\ \dots, \text{n} \tag{20}$$

where the conditional variance of each ε_i is constant and equal to σ^2,or:

$$var(\varepsilon|X) = \sigma^2 I, \tag{21}$$

if considered the vector of the errors ε, where I is the N identity matrix and $\sigma^2 I$ is the error's variance-covariance matrix.

There are some reasons as to why heteroskedasticity happens. One of the reasons is measurement errors. The condition that the variables are measured without error is not always possible to fulfill. Other reasons that can raise heteroskedasticity problems are the skewness in the distribution of one or more explanatory variables and the incorrect data transformation of the model. Both when there is an increase of the independent variable and that results in an increase in the variance of the errors, and when the errors of an independent variable range from the extremes positive to negative, heteroskedasticity can also be expected.

The heteroskedastic errors lead to some consequences, for example, it makes the estimators for the standard errors biased and inconsistent, which doesn't allow for the use of t and F tests as it leads to bias in statistics and confidant levels. Another consequence of heteroskedasticity is that the OLS estimator is no longer BLUE and thus the estimators are less efficient. However, as the properties of the OLS estimators depend upon two other assumptions that may not be violated, we can consider the estimators still consistent and unbiased (Berry and Feldman, 1985).

As we are dealing with panel data, we will be testing for heteroskedasticity in ε_{it}, using the fixed effects residuals $\hat{\varepsilon}_{it}$. For this, we will use a variant of the Breusch Pagan test. The auxiliary regression of the test regresses $\hat{\varepsilon}_{it}^2$ upon a constant and K variables z_{it}, that may cause heteroskedasticity. The alternative hypothesis is given by

$$\sigma_i^2 = \sigma^2 h(z_{it}'\alpha), \tag{22}$$

where h is an unknown continuously differentiable function with $h(0) = 1$ (Verbeek, 2012). The null hypothesis is H_0: $\alpha = 0$. Under this hypothesis, the test statistic is N(T-1) times R^2 of the auxiliary regression. This test has an asymptotic Chi-squared distribution with j degrees of freedom. Heteroscedasticity is present in the case that we reject the null hypothesis. In that case, H_1: $\alpha \neq 0$, which means we should revise the specification of our model or find an alternative estimator in order to solve the heteroscedasticity problem.

3.8. Unit Root

With the advance in economic studies, panel data can, nowadays, be used to deal with both long time series observations (i) and large number of groups of individuals (t), which is very convenient. However, this model isn't free from complications. In panel data, the distribution of the variable of interest shouldn't depend upon time, that is, the data generating process should be stationary. In the cases this doesn't happen, we are in the presence of non-stationarity and one of the biggest causes for this are unit roots. One of the main consequences of this is a spurious regression, which can confuse the interpretation of the estimated results, causing misinterpretations (Breitung & Pesaran, 2008).

With this said, it is not adequate to use conventional methods of analysis of unit roots in panel data with large i and t, since these conventional models are based on a large i and a small t, as it used to be when panel data was first developed (Hsiao, 2007).

One of the tests that can be used is the Im Pesaran and Shin test (IPS), which is a standardized average of individual Augmented Dickey–Fuller (ADF) statistics and follows a normal distribution. The null hypothesis in this test is that all the series included have unit root, which means that they are not stationary. Rejecting this hypothesis means that there is at least one series that is stationary. Therefore, the alternative hypothesis states that some of the series included in the panel are stationary.

Another test that can be used is the Maddala and Wu test (MW test) which is inspired by a Fisher type test that combines p-values from unit root tests for each cross-section i. The procedure recommended by this test is nonparametric and doesn't require a balanced panel. The way this test works is that it conducts unit root tests for each time series individually and then combines the p-values of those teste in order to generate an overall test (Breitung & Pesaran, 2008).

Does the number of Robinhood’s subscribers influence the stock prices?

Does the number of Robinhood's subscribers influence the stock prices?

4. Hypothesis and Empirical Model

Hypotheses 1

We expect that the increase in the number of subscribers is positively related to the stock prices, that is, an increase in the number of subscribers will lead to an increase in the stock prices.

Hypotheses 2

We expect that the relation expressed in Hypothesis 1 is much stronger since the COVID-19 lockdown.

4.1. Empirical Model

The focus of this study is to verify whether changes in the number of Robinhood's subscribers impact stocks returns. For this we will be using the following multiple linear regression model:

$$Stock\ Return_{it} = \alpha_i + \beta_1 \text{User Holding Return}_{it} + \beta_2 \text{Market Return}_t + \varepsilon_{it}, \qquad (23)$$
$$i = 1, 2, \ldots, 20 \text{ and } t = 1, 2, \ldots, 564$$

where:

$Stock\ Return_{it}$ - the log of the return of the stock *i* at time *t*;

α_i - the intercept regression coefficient;

β_1 - the regression coefficient associated with changes in the log return of the number of subscriptions in the stock *i*;

$\text{User Holding Return}_{it}$ - the log return of the number of subscriptions in the stock *i*;

β_2 - the regression coefficient associated with changes in the log return of the S&P 500 index at time *t*;

Market Return_t - the log return of the S&P 500 index at time *t*;

ε_{it} - the regression error for stock *i* at time *t*;

Does the number of Robinhood’s subscribers influence the stock prices?

Does the number of Robinhood's subscribers influence the stock prices?

5. Data and Empirical Results

In this section, we will present the data that was used in our study and explain the transformations that had to be made in order to concretize this study. Secondly, we will estimate the models presented in the previous section: the model using all the data, the model using only pre-lockdown data and the model using post-lockdown data. Lastly, we will interpret the obtained results and discuss possible economical and financial reasons for these.

5.1. Data

To be able to achieve the goal of this thesis, we will be using the number of subscribers of Robinhood for each stock and the respective adjusted closing daily price. The number of subscribers of Robinhood for each stock indicates the number of people that invest on a specific stock that are Robinhood's client. To verify the veracity of our first hypothesis, we use the S&P 500 too. This index is composed by the largest 500 publicly listed companies in the United States of America. The S&P 500 will work as a representative of the market, as it not only is the best benchmark of the American stock market, but also a statistical measure of the American economy. This will allow the observation of how much the market influences the prices and infer if the Robinhood's subscribers effect is relevant or not.

An independent website called Robintrack (www.robintrack.net) gathered data from Robinhood between May of 2018 and August 2020. Therefore, the period we are working with is between 02/05/2018 and 13/08/2020, since our data was extracted from Robintrack. The available data on this site is on an hourly basis. The adjusted closing price is available at www.finance.yahoo.com, on a daily basis and the S&P 500 is available in Bloomberg, also on a daily basis.

We have chosen the top 20 companies that belong to S&P 500 and have the biggest market capitalization. However, we had to switch Berkshire Hathaway (which belongs to this list) by Comcast, that is the 21^{st} biggest market capitalization in S&P 500, because there was a scarcity of data on the number of subscribers for Berkshire Hathaway. We have 10 companies that belong to NASDAQ and 10 that belong to NYSE as we can see in the table below.

Does the number of Robinhood's subscribers influence the stock prices?

Table 5.1 List of stocks

S&P 500	
NASDAQ	**NYSE**
Apple (AAPL)	JPMorgan Chase (JPM)
Microsoft (MSFT)	Visa (V)
Amazon (AMZN)	Johnson & Johnson (JNJ)
Alphabet (GOOG)	Walmart (WMT)
Facebook (FB)	Disney (DIS)
Tesla (TSLA)	Mastercard (MA)
Nvidia (NVDA)	Procter & Gamble (PG)
Paypal (PYPL)	UnitedHealth Group (UNH)
Intel (INTC)	Bank of America (BAC)
Comcast (CMCSA)	The Home Depot (HD)

Due to the fact that only the closing price of the stock and the closing points of the S&P 500 are available, this dissertation will work with daily data. This means that the data of the number of Robinhood's subscribers has to be rearranged. To do that, we keep only the last entry of the day, that is, the closing value of Robinhood's subscribers.

5.2. Empirical Results

After joining all the 20 companies' data together, we introduce it in R and define it as panel data. Afterwards, we test the stationarity of the data. This is done by testing the existence of a unit root based on the Im Pesaran and Shin test (IPS). This test is a standardized average of individual Augmented Dickey–Fuller (ADF) statistics and follows a normal distribution. The null hypothesis in the IPS unit root test stated that all the series included have unit root or in other words, are non-stationary. While, on the other hand, alternative hypotheses stated that some of the series included in the panel are stationary. Thus, rejection of null means that there is at least one series that is stationary. The results are in table 5.2:

Does the number of Robinhood's subscribers influence the stock prices?

Table 5.2 IPS test results

Variable	*Statistics*	*P-value*
Stock Prices	1.547	(0.939)
User Holding	19.884	(1)
Market	-1.579	(0.057)

From the results obtained, we conclude that all variables are non-stationary with a 5% significance level. However, the test result for the variable "Market" fell a little short, so we decided to do the Maddala-Wu unit root test, which has the same null and alternative hypotheses of stationarity of IPS. Moreover, in parallel, we observe the graph representation of the time series to confirm the result.

Table 5.3 Maddala-Wu test result

	Statistics	*P-value*
Market	39.6	(0.488)

Graph 5.1 S&P 500 time series

Based on these results, the conclusion stated previously is right, that is, the variable "Market" has a unit root, therefore is non-stationary.

In order to proceed with our analyses, we must make changes in the data to eliminate the unit root and, with this, make the variables stationary. This is done by calculating the daily return of logarithm of all the variables, making their variance and mean more constant. This way, we hope to guarantee the stationarity of all the variables.

In order to verify if the changes made previously were enough to make the variables stationary, we repeat the IPS test:

Table 5.4 IPS test results for the log return of the variables

Variable	*Statisitics*	*P-value*
Stock Return	-117.407	(0)
User Holding Return	-57.018	(0)
Market Return	-58.148	(0)

The results of the tests (table 5.4) prove that all variables are stationary since the p-value < 0.05.

Afterwards, we test which model is more appropriate for our data. To do that, we are going to run 3 tests: F, Breusch-Pagan and Hausman tests. The F test is meant to discriminate between the fixed effect model and the pooled OLS. The null hypothesis of this test is that the pooled OLS is the most adequate model to fit the data and the alternative hypothesis states that it is the fixed effect model. The Breusch-Pagan Lagrange Multiplier test is used to choose between the random effect model and the pooled OLS. The null hypothesis of this test is that the pooled OLS is the most adequate model to use and the alternative hypothesis states that it is the random effect model. Although our T is large and therefore it's indifferent to use the random or the fixed effect model, we use the Hausman test to choose between these models. The null hypothesis of the Hausman test is that the random effect model is more appropriate, and the alternative hypothesis is that is the fixed effect model. The results of these tests are expressed in the table below:

Table 5.5 Results of the tests to choose the model

Test	*Statistics*	*P-value*
F	1.6358	(0.03976)
Breusch-Pagan Lagrange Multiplier	3.0084	(0.08284)
Hausman	14.939	(0.0005701)

By analyzing the tests results, we conclude that the most adequate model seems to be the fixed effect model, since we reject the null hypothesis on both F and Hausman tests (p-value < 0.05).

Does the number of Robinhood's subscribers influence the stock prices?

Afterwards, we are going to check if there is any autocorrelation in the error terms. The test that is going to be used for that purpose is the modified Durbin-Watson test. The null hypothesis of this test is that there is no autocorrelation in the error terms and the alternative hypothesis is that there is. The obtained results are:

Table 5.6 Durbin Watson test results

	Statistics	*P- value*
Durbin Watson	2.0115	(0.7311)

The result of the test demonstrates that there is no autocorrelation in the error terms since the p-value is higher than 0.05. However, we decide to compute a Breusch-Godfrey test, with lag = 1, to confirm the result. The following results were obtained:

Table 5.7 Breusch-Godfrey test result

	Statistics	*P- value*
Breusch-Godfrey	0.3768	(0.5393)

The null hypothesis is not rejected, since the p-value is higher than 0.05, which confirms the conclusion of the Durbin-Watson test that there is no autocorrelation in the error terms.

Finally, we test the heteroscedasticity of errors in our model with the Breusch-Pagan test, whose null hypothesis is that the errors are homoscedastic. The results we obtained are:

Table 5.8 Homoscedasticity test result

	Statistics	*P- value*
Breusch-Pagan	5815.5	(< 2.2e-16)

Since the p-value is lower than 0.05, the null hypothesis is rejected. We can, then, conclude that the errors are heteroskedastic. As a consequence of this result, we need to compute the robust standard errors to obtain reliable estimates for our model. To do that, we use the White heteroskedasticity-consistent covariance estimator.

After computing all the diagnostic tests to our models and computing the robust standard errors to deal with heteroskedasticity problems, the estimates for our explanatory variables are the following:

Does the number of Robinhood's subscribers influence the stock prices?

Table 5.9 Estimates of the model

Variables	*Estimate*	*Std. Error*	*P- value*
User Holding Return	-0.385512	0.065606	(4.318e-09)
Market Return	1.057957	0.019050	(< 2.2e-16)

The results from table 5.9 reveal that both estimates from the coefficients are statistically significant (p-value < 0.05). Given these results, we can conclude that there is a negative relation between the return of the number of subscribers from Robinhood and the stock returns, that is, if the return of the number of Robinhood's subscribers increase 1%, the predicted return of the stock will decrease 0.3855%, supposing all the rest constant. The estimate of the Market Return is near 1, meaning almost a perfect relation between the Market Return and the Stock Return.

These results go against our first hypothesis which states that the return of the number of subscribers are positively related to the return on stocks, at least for those included in our sample. Now, we are going to test out our second hypothesis and split our sample in two: data before lockdown and data after lockdown in America.

Table 5.10 Timeline data after split

Before lockdown	02/05/2018 to 18/03/2020
After lockdown	19/03/2020 to 13/08/2020

In order to check if the variables are still stationary after the split of the data, we do the IPS test again. The results of the tests are at the table below:

Table 5.11 IPS test results for the log return of the variables (split data)

Variable	*Period*	*Statisitics*	*P-value*
Stock Return	Before lockdown	-107.92	(0)
	After lockdown	-54.07	(0)
User Holding Return	Before lockdown	-58.402	(0)
	After lockdown	-29.596	(0)
Market Return	Before lockdown	-43.78	(0)
	After lockdown	-62.623	(0)

With these results, we can conclude that all the variables are still stationary since we reject the null hypothesis (p-value < 0.05).

Next, we test which model is more appropriate for our data. The results of the tests are expressed in table 5.12:

Table 5.12 Results of the tests to choose the more suitable model (split data)

Test	*Period*	*Statistics*	*P-value*
F	Before lockdown	1.071	(0.3741)
	After lockdown	2.2723	(0.001338)
Breusch-Pagan Lagrange Multiplier	Before lockdown	0.0016089	(0.968)
	After lockdown	12.518	(0.000403)
Hausman	Before lockdown	10.558	(0.005098)
	After lockdown	6.5676	(0.03749)

By the tests results in table 5.12, we can conclude that the model that seems to be more adequate for the period before lockdown is the pooled OLS model because in both, the F test and the LM Breusch-Pagan test, the null hypothesis is not rejected. The model that is more adequate for the period after lockdown is the fixed effects model, since the null hypothesis is rejected in F, Breusch-Pagan and Hausman tests.

Afterwards, we check whether there is autocorrelation in the error terms. For this we do both the modified Durbin Watson test and the Breusch-Godfrey test with lag = 1. The obtained results are (see table 5.13):

Table 5.13 Autocorrelation tests results (split data)

	Period	*Statistics*	*P- value*
Durbin Watson	Before lockdown	2.0297	(0.9231)
	After lockdown	1.9938	(0.452)
Breusch-Godfrey	Before lockdown	2.7591	(0.0967)
	After lockdown	0.016447	(0.898)

With the p-value of both tests being higher than 0.05, we confirm that there is no autocorrelation in the error terms.

Finally, we test the heteroscedasticity of errors in our model with the Breusch-Pagan test. The results we obtained are (see table 5.14):

Does the number of Robinhood's subscribers influence the stock prices?

Table 5.14 Homoscedasticity test result (split data)

	Period	*Statistics*	*P- value*
Breusch-Pagan	Before lockdown	2381.4	(< 2.2e-16)
	After lockdown	1052.5	(< 2.2e-16)

Due to the fact that the null hypothesis is rejected in both periods, we can conclude that the errors are heteroskedastic in these periods. As a consequence of this result, we need to compute the robust standard errors for each period, using once again the White heteroskedasticity-consistent covariance estimator, to obtain reliable estimates for both models.

After computing all the necessary tests to our models and computing the robust standard errors to deal with heteroskedasticity problems, the estimates for our explanatory variables are the following:

Table 5.15 Estimates of the model (split data)

Variables	*Period*	*Estimate*	*Std. Error*	*P- value*
(Intercept)	Before lockdown	0.00114818	0.00016989	(1.482e-11)
User Holding Return	Before lockdown	-0.50017548	0.07119336	(2.283e-12)
	After lockdown	-0.23254	0.12195	(0.05667)
Market Return	Before lockdown	1.08657231	0.02415493	(< 2.2e-16)
	After lockdown	0.98164	0.019050	(< 2.2e-16)

The results from table 5.15 reveal that all the estimates for the coefficients are statistically significant (p-value < 0.1). Given the results, we can conclude, once again, that our hypothesis is not verified. There is still a negative relation between the return of the number of subscribers from Robinhood and the stock returns, although the impact is less negative. This time, if the return of the number of subscribers of Robinhood increase by 1%, the predicted return of the stock will decrease 0.23254%, supposing all the rest constant. Again, the estimate of the Market Return is near 1, which indicates a close to perfect relation between the Market Return and the Stock Return.

Does the number of Robinhood’s subscribers influence the stock prices?

6. Conclusion

The factors that move the stock prices has always been an intriguing question. Knowing the reason behind their rises and declines would lead us into predictions about what would happen in the future. These forecasts are not only extremely important to companies listed on the stock exchange market, as they could adjust their management strategy considering the behavior of the market and the market's reaction to certain decisions arising from the company's management, but also important for investors, as they too could take advantages of the predictions to sell or buy at the right times. However, this issue is not linear, because if there were to be a soul that could accurately predict market shifts, the news would quickly spread, everybody would then do it and the effect would, of course, disappear. This is what is called an efficient stock market.

In this study, we were searching to know whether the retail investors, particularly from Robinhood, have had an influence on stock prices. Our results suggest that they didn’t. Despite the growing number of Robinhood’s retail investors, the easy access to financial markets nowadays and the simplified way the broker presents their services, the influence is not statistically sufficient. Better said, at least it doesn't influence the group of 20 stocks that have been chosen in this present work, the ones with the highest market capitalization of S&P 500. They were actually chosen due to the fact that they have the highest market capitalization, therefore these stocks predict the market better than any others. Even during the COVID-19 lockdown, the relationship between stock prices and the number of Robinhood subscribers was negative. This result goes against our expectations because with the lockdown, there was a huge increase in subscribers, therefore an increase in stock prices caused by them was expected.

Stein’s similar study (Stein, 2020) about the 5 biggest factors that contribute to stocks entering the “100 Most Popular” list on Robinhood, is in accordance with our findings. Stein (2020) noticed that when stocks entered this list, they received a lot of investor’s attention, which induced a rise of subscribers on those stocks and, consequently, a rise in their prices. However, this effect has shown to be short lived, and as soon as the novelty effect disappeared, stock prices decreased again. Moreover, Stein (2020) also noticed that this phenomenon was clearly stronger for smaller stocks. Since our study was made with 20 of the most influential stocks on the market, the effect depicted by the author may not have occurred.

With this said, we can conclude that the retail investors of Robinhood don’t yet have a sufficiently considerable volume of trading to be able to influence the market. At the end, we are still talking about retail investors. Despite this, we speculate that the tendency is in the direction of our hypothesis, that is, retail investors will become more and more influential throughout time. We expect this because of the easy access to the platform and app, as well as to knowledge on the subject. Not only that but the negative interest rates encourage the search for more profitable ways of investment.

This study, as all studies, entails some limitations. One of them is the stock choice criteria. We have chosen the biggest 20 stocks, but as stated before, stock size may play an influence in the relation between number of subscribers and stock prices. Had the criterion been other, the results could have been different. Another limitation is one that neighbors linear regression methodologies. In our model, it is assumed that the increase in the number of subscribers on one stock will impact the prices of that stock in the same way that the same increase of the number of subscribers on another stock would impact that stock.

This dissertation is relevant in a way that it elucidates about the influence of the number of Robinhood’s subscribers on the market, thus helping those who invest in the market. On the other hand, it is also purposeful from a researcher’s point of view, in that it investigates the effects of the demand on the market. For future studies, we suggest a remake of the study with different stocks, namely smaller ones, in order to investigate their effect on stock prices.

Does the number of Robinhood's subscribers influence the stock prices?

7. References

Abudy, M. M. (2020). Retail investors' trading and stock market liquidity. *The North American Journalof Economics and Finance, 54,* 101281.

Aouadi, A., Arouri, M., & Teulon, F. (2013). Investor attention and stock market activity: evidence from france. *Economic Modelling, 35,* 674-681.

Baltagi, B. (2005). *Econometric analysis of panel data* (3rd ed.). John Wiley & Sons Ltd.

Barber, B. M., Huang, X., Odean, T., & Schwarz, C. (2020). *Attention induced trading and returns: evidence from robinhood users*. SSRN 3715077. https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3715077.

Barber, B.M., Odean, T. (2008). All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies, 21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079.

Barber, B. M., Odean, T., & Zhu, N. (2008). Do retail trades move markets? *The Review of Financial Studies*, *22*(1), 151-186.

Breitung, J. & Pesaran, M. H. (2005). Unit roots and cointegration in panels. *Deutsche Bundesbank*. https://www.econstor.eu/bitstream/10419/19627/1/200542dkp.pdf.

Berry, W. D. & Feldman, S. (1985). *Multiple regression in practice*. SAGE Publications.

Cameron, A. C. & Trivedi, P. K. (2010). *Microeconometrics Using Stata*. Stata Press.

Cao, H. H., Han, B., Hirshleifer, D., & Zhang, H. H. (2011). Fear of the unknown: Familiarity and economic decisions. *Review of finance, 15*(1), 173-206.

Ding, R., & Hou, W. (2015). Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions and Money, 37*, 12-26.

Glossner, S., Matos, P., Ramelli, S., & Wagner, A. F. (2020). *Where do institutional investors seek shelter when disaster strikes? Evidence from COVID-19*. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3661413.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251-71.

Hsiao, C. (2007). Panel data analysis—advantages and challenges. *Test*, 16(1), 1-22.

Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3663970.

Kenkel, B. (2021). *Practical data analysis for political scientists*. Vanderbilt University. https://bkenkel.com/pdaps/.

Liaukonytė, J., & Žaldokas, A. (2019). Background noise? TV advertising affects real time investor behavior. *SSRN*. https://doi.org/10.2139/ssrn.3336176.

Mazur, M., Dang, M., & Vega, M. (2020). COVID-19 and the March 2020 Stock Market Crash. Evidence from S&P1500. *Finance Research Letters, 38*, 101690. https://doi.org/10.1016/j.frl.2020.101690.

Merton, R. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance, 42*, 483–510. https://doi.org/10.1111/j.1540-6261.1987.tb04565.x.

Ozik, G., Sadka, R., & Shen, S. (2020). Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. *SSRN*.

Pagano, M. S., Sedunov, J., & Velthuis, R. (2020). How did retail investors respond to the COVID-19 pandemic? the effect of robinhood brokerage customers on market quality. *Finance Research Letters, 101946.*

Peress, J., & Schmidt, D. (2020). Glued to the TV: Distracted noise traders and stock market liquidity. *The Journal of Finance, 75*(2), 1083-1133.

Robinhood (2021, September 24). https://robinhood.com/us/en/.

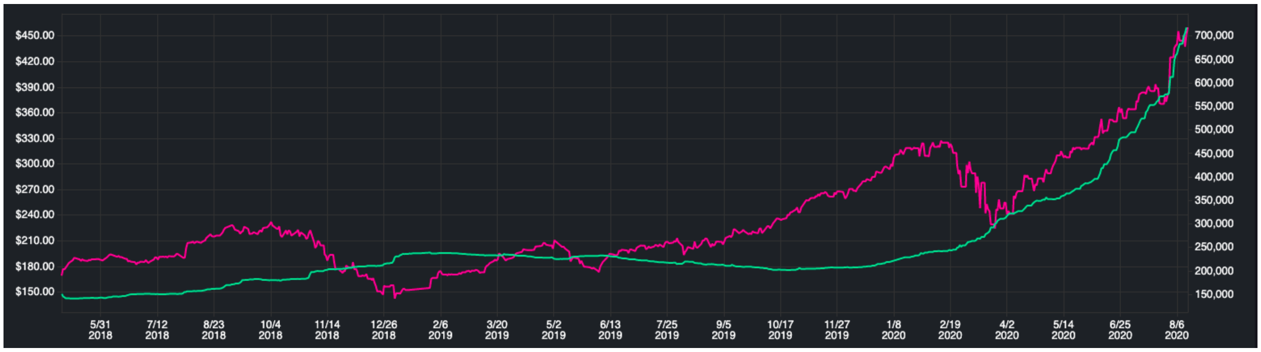
Robintrack (2020, august 23). http://ww.robintrack.net.
Rooney, K. (2020, November 23). Fintech app robinhood is driving a retail trading renaissance during stock market's wild ride. *CNBC*. https://www.cnbc.com/2020/06/17/robinhood-drives-retail-trading-renaissance-during-markets-wild-ride.html#close.
Schmidheiny, K. (2012). *Panel Data: Fixed and Random Effects*. Short Guides to Microeconometrics. https://www.schmidheiny.name/teaching/panel2up.pdf.
Stein, R. (2020). *The Top 5 predictable effects of new entries in Robinhood's "100 Most Popular" list*. SSRN 3694588. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3694588.
Verbeek, M. (2004). *A guide to modern econometrics* (2nd ed.). John Wiley & Sons Ltd.
Verbeek, M. (2012). *A guide to modern econometrics* (4th ed.). John Wiley & Sons Ltd.
Welch, I. (2020). Retail raw: wisdom of the robinhood crowd and the COVID crisis. *National Bureau of Economic Research*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3696066.
Wooldridge, J. M. (2017). Introductory Econometrics A Modern Approach. In Economica (6th ed., Vol. 42, Issue 165). https://doi.org/10.2307/2553001

Does the number of Robinhood's subscribers influence the stock prices?

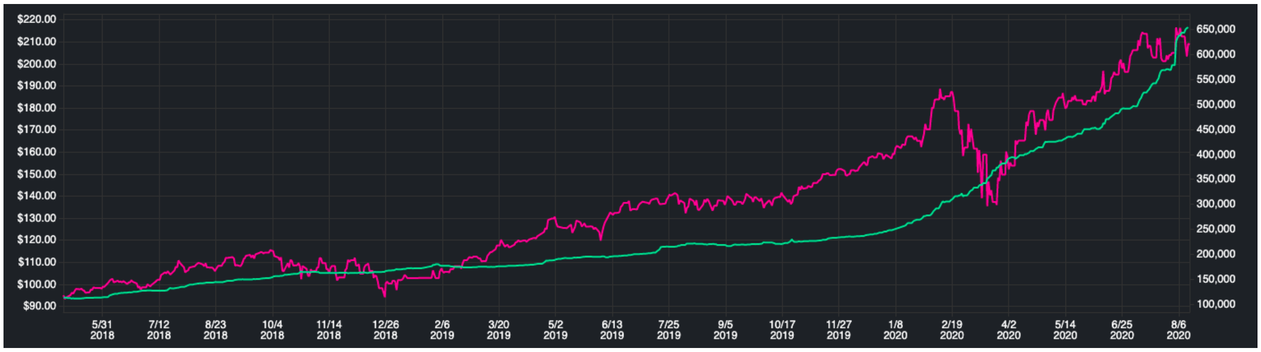
8. Annexes

Annex A Graphic representation of stock prices and the number of Robinhood's subscribers of that stock

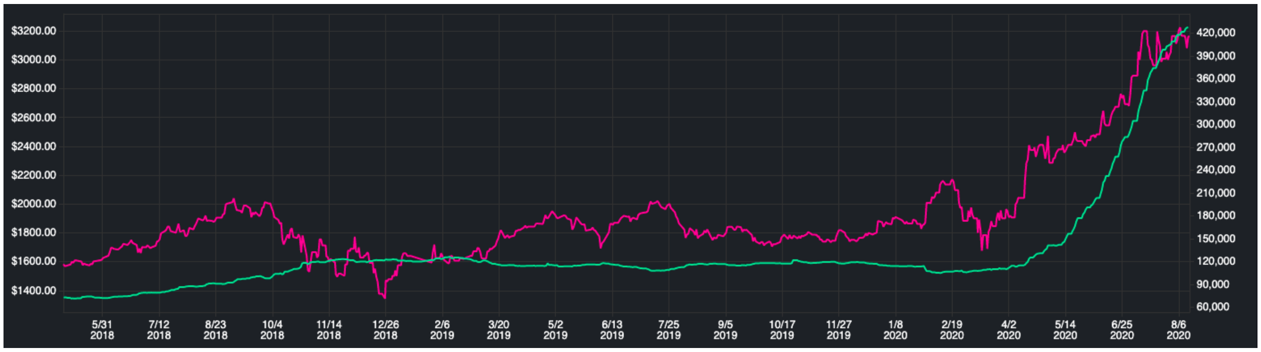
1) Apple



2) Microsoft



3) Amazon

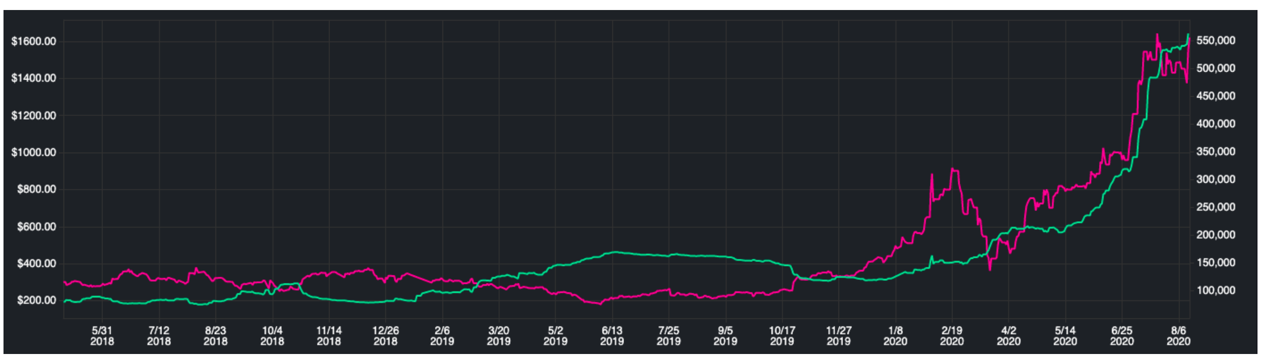


Does the number of Robinhood’s subscribers influence the stock prices?

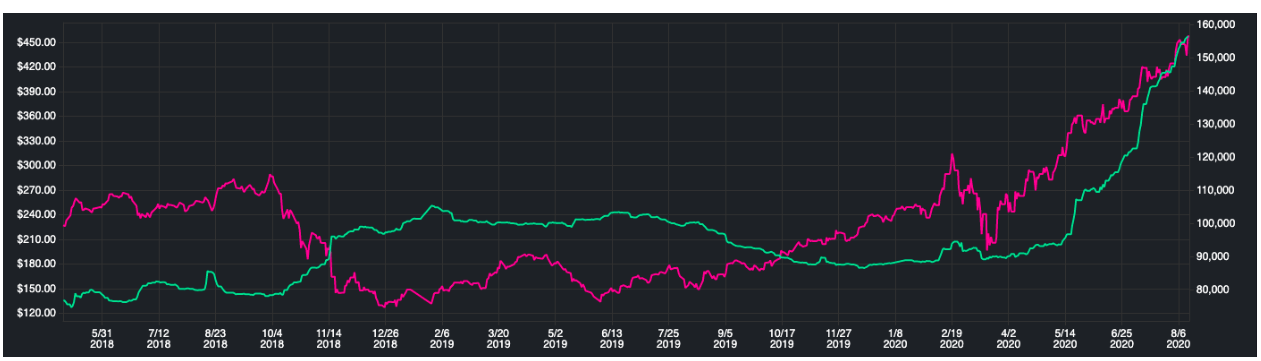
4) Alphabet

5) Facebook

6) Tesla

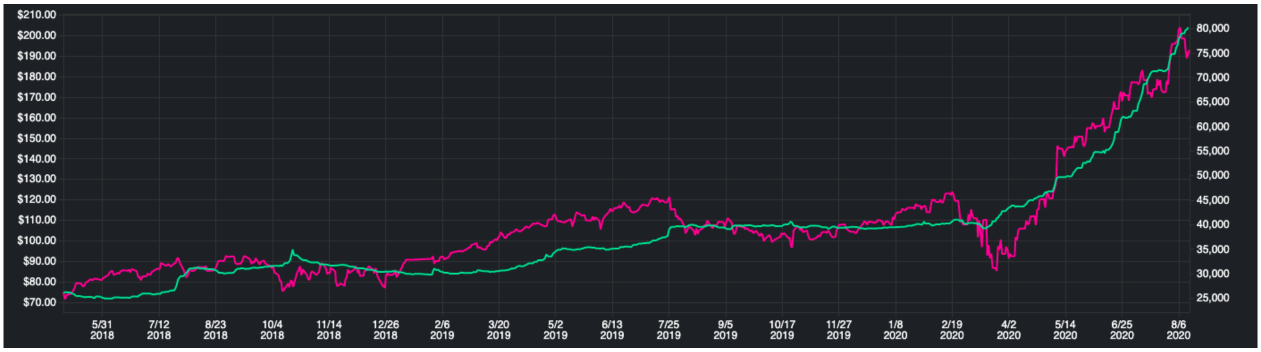


7) Nvidia

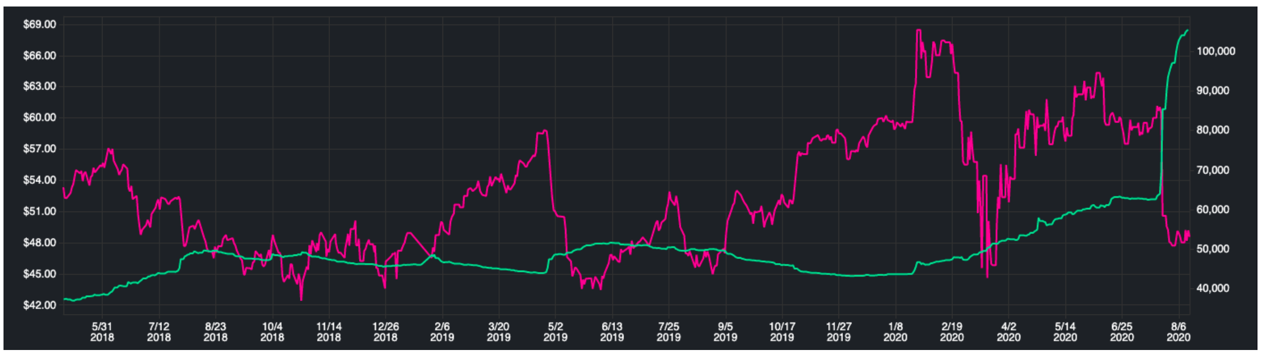


Does the number of Robinhood's subscribers influence the stock prices?

8) Paypal



9) Intel

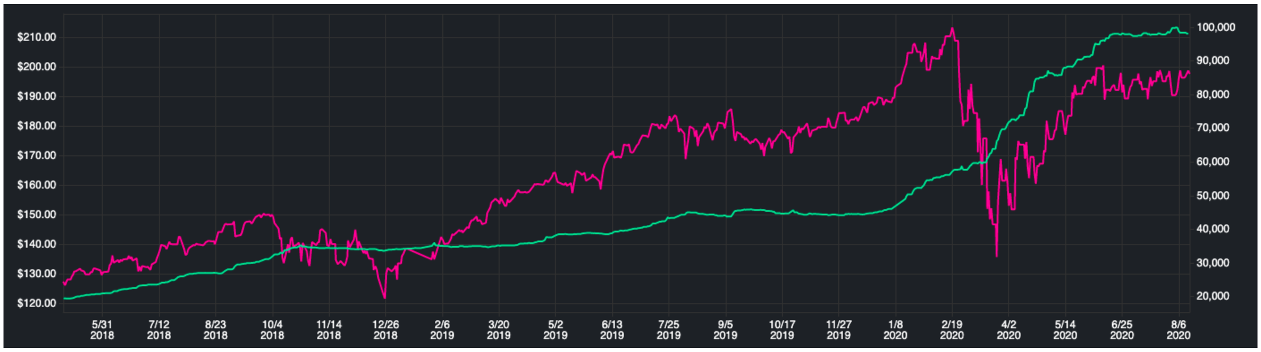


10) Comcast

11) JPMorgan Chase

Does the number of Robinhood's subscribers influence the stock prices?

12) Visa



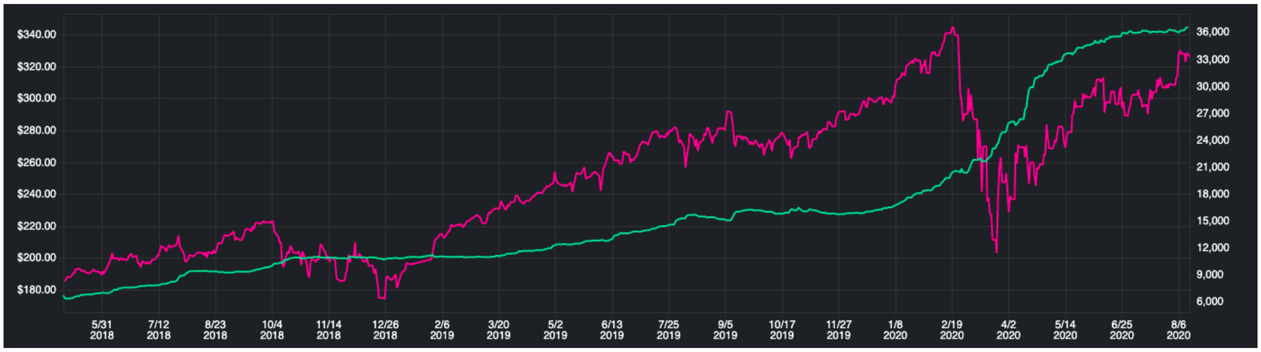
13) Johnson & Johnson

14) Walmart

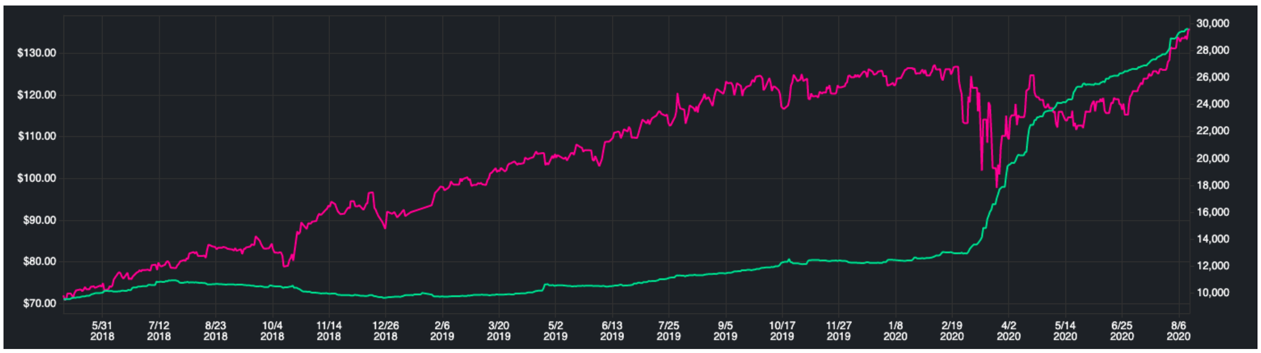
15) Disney

Does the number of Robinhood's subscribers influence the stock prices?

16) Mastercard

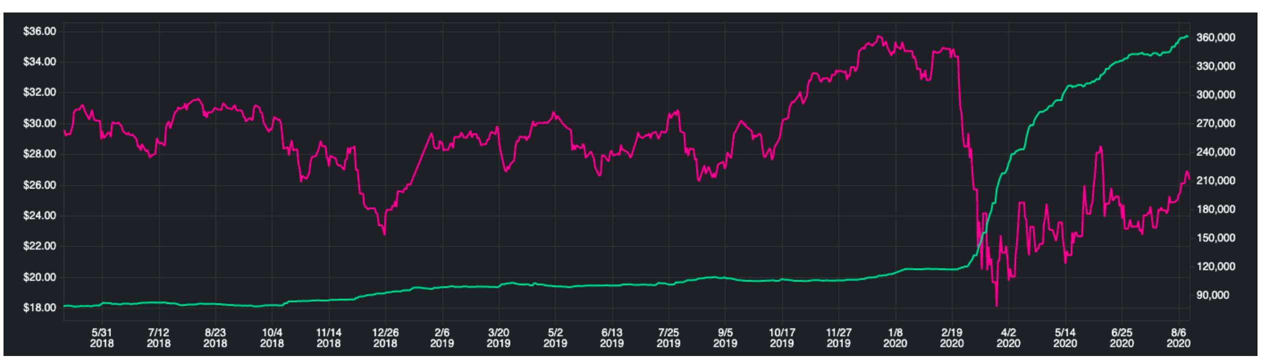


17) Procter & Gamble



18) UnitedHealth Group

19) Bank of America



Does the number of Robinhood's subscribers influence the stock prices?

20) The Home Depot

Annex B Output of unit root test

1) Stock Prices

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 4 lags (max: 5)
statistic (Wtbar): 1.547
p-value: 0.939
```

2) User Holdings

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 5 lags (max: 5)
statistic (Wtbar): 19.884
p-value: 1
```

3) Market

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 4 - 4 lags (max: 5)
statistic (Wtbar): -1.579
p-value: 0.057

Maddala-Wu Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 4 - 4 lags (max: 5)
statistic: 39.6
p-value: 0.488
```

4) Stock Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 5 lags (max: 5)
statistic (Wtbar): -117.407
p-value: 0
```

5) User Holding Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 4 lags (max: 5)
statistic (Wtbar): -57.018
p-value: 0
```

6) Market Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 3 - 3 lags (max: 5)
statistic (Wtbar): -58.148
```

Does the number of Robinhood’s subscribers influence the stock prices?

```
p-value: 0
```

Annex C Output for the pooling model

```
Pooling Model

Call:
plm(formula = `Stock return` ~ `Log(diff user)` + `Market return`,
    data = dados.p, model = "pooling")

Balanced Panel: n = 20, T = 563, N = 11260

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-0.23908779 -0.00732807 -0.00063623  0.00630409  0.20759559

Coefficients:
                    Estimate  Std. Error  t-value  Pr(>|t|)
(Intercept)       0.00138840  0.00015941   8.7095 < 2.2e-16 ***
`Log(diff user)` -0.38070988  0.01555369 -24.4771 < 2.2e-16 ***
`Market return`   1.05800605  0.00978563 108.1183 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    6.3885
Residual Sum of Squares: 3.0428
R-Squared:      0.5237
Adj. R-Squared: 0.52362
F-statistic: 6188.69 on 2 and 11257 DF, p-value: < 2.22e-16
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex D Output for the fixed effects model

```
Oneway (individual) effect Within Model

Call:
plm(formula = `Stock return` ~ `Log(diff user)` + `Market return`,
    data = dados.p, model = "within")

Balanced Panel: n = 20, T = 563, N = 11260

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-0.24202469 -0.00730114 -0.00056216  0.00630310  0.20557420

Coefficients:
                   Estimate Std. Error t-value  Pr(>|t|)
`Log(diff user)` -0.3855119  0.0156027 -24.708 < 2.2e-16 ***
`Market return`   1.0579573  0.0097804 108.171 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    6.383
Residual Sum of Squares: 3.0344
R-Squared:      0.52461
Adj. R-Squared: 0.52372
F-statistic: 6200.69 on 2 and 11238 DF, p-value: < 2.22e-16
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex E Output for the random effects model

```
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = `Stock return` ~ `Log(diff user)` + `Market return`,
    data = dados.p, model = "random")

Balanced Panel: n = 20, T = 563, N = 11260

Effects:
                    var   std.dev share
idiosyncratic 2.700e-04 1.643e-02     1
individual    4.362e-09 6.604e-05     0
theta: 0.004517

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-0.23910317 -0.00732812 -0.00063774  0.00630196  0.20758840

Coefficients:
                    Estimate  Std. Error  z-value  Pr(>|z|)
(Intercept)       0.00138850  0.00016009   8.6731 < 2.2e-16 ***
`Log(diff user)` -0.38075284  0.01555401 -24.4794 < 2.2e-16 ***
`Market return`   1.05800561  0.00978551 108.1196 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    6.3884
Residual Sum of Squares: 3.0427
R-Squared:      0.52371
Adj. R-Squared: 0.52363
Chisq: 12377.8 on 2 DF, p-value: < 2.22e-16
```

Does the number of Robinhood's subscribers influence the stock prices?

Does the number of Robinhood's subscribers influence the stock prices?

Annex F Output of F test

```
      F test for individual effects

data:  `Stock return` ~ `Log(diff user)` + `Market return`
F = 1.6358, df1 = 19, df2 = 11238, p-value = 0.03976
alternative hypothesis: significant effects
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex G Output of Breusch-Pagan Lagrange Multiplier test

```
      Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels

data:  `Stock return` ~ `Log(diff user)` + `Market return`
chisq = 3.0084, df = 1, p-value = 0.08284
alternative hypothesis: significant effects
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex H Output of Hausman test

```
      Hausman Test

data:  `Stock return` ~ `Log(diff user)` + `Market return`
chisq = 14.939, df = 2, p-value = 0.0005701
alternative hypothesis: one model is inconsistent
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex I Output of the autocorrelation test

1) Durbin-Watson test for serial correlation in panel models

```
data:  `Stock return` ~ `Log(diff user)` + `Market return`
DW = 2.0115, p-value = 0.7311
alternative hypothesis: serial correlation in idiosyncratic errors
```

2) Breusch-Godfrey/Wooldridge test for serial correlation in panel models

```
data:  `Stock return` ~ `Log(diff user)` + `Market return`
chisq = 0.3768, df = 1, p-value = 0.5393
alternative hypothesis: serial correlation in idiosyncratic errors
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex J Output of the homoscedasticity test

```
Breusch-Pagan test

data:  fixedeff
BP = 5815.5, df = 2, p-value < 2.2e-16
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex K Output for the fixed effects model after controlling for the heteroscedasticity in the model

```
t test of coefficients:

                   Estimate Std. Error t value  Pr(>|t|)    
`Log(diff user)` -0.385512   0.065606 -5.8762 4.318e-09 ***
`Market return`   1.057957   0.019050 55.5345 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex L Output of unit root test after lockdown

1) Stock Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 5 lags (max: 5)
statistic (Wtbar): -54.07
p-value: 0
```

2) User Holding Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 3 lags (max: 5)
statistic (Wtbar): -29.596
p-value: 0
```

3) Market Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 0 lags (max: 5)
statistic (Wtbar): -62.623
p-value: 0
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex M Output for the pooling model after lockdown

```
Pooling Model

Call:
plm(formula = `Stock Return after` ~ `Log(diff user) after` +
    `Market return after`, data = dados.p2, model = "pooling")

Balanced Panel: n = 20, T = 102, N = 2040

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.1295492 -0.0116198 -0.0012716  0.0093174  0.1369195

Coefficients:
                        Estimate  Std. Error t-value  Pr(>|t|)
(Intercept)            0.00220854  0.00053506  4.1277 3.812e-05 ***
`Log(diff user) after` -0.20756332  0.03587478 -5.7858 8.337e-09 ***
`Market return after`   0.98267075  0.02275847 43.1782 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1.7872
Residual Sum of Squares: 0.91579
R-Squared:      0.48759
Adj. R-Squared: 0.48709
F-statistic: 969.184 on 2 and 2037 DF, p-value: < 2.22e-16
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex N Output for the fixed effects model after lockdown

```
Call:
plm(formula = `Stock Return after` ~ `Log(diff user) after` +
    `Market return after`, data = dados.p2, model = "within")

Balanced Panel: n = 20, T = 102, N = 2040

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.1211838 -0.0114442 -0.0010670  0.0097601  0.1269629

Coefficients:
                        Estimate Std. Error t-value  Pr(>|t|)
`Log(diff user) after` -0.232543   0.036483  -6.374 2.275e-10 ***
`Market return after`   0.981644   0.022627  43.384 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1.771
Residual Sum of Squares: 0.8966
R-Squared:      0.49374
Adj. R-Squared: 0.48848
F-statistic: 984.064 on 2 and 2018 DF, p-value: < 2.22e-16
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex O Output for the random effects model after lockdown

```
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = `Stock Return after` ~ `Log(diff user) after` +
    `Market return after`, data = dados.p2, model = "random")

Balanced Panel: n = 20, T = 102, N = 2040

Effects:
                   var   std.dev share
idiosyncratic 4.443e-04 2.108e-02 0.992
individual    3.537e-06 1.881e-03 0.008
theta: 0.2571

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.1264808 -0.0115038 -0.0011819  0.0094245  0.1343765

Coefficients:
                          Estimate  Std. Error z-value  Pr(>|z|)
(Intercept)             0.00228324  0.00067978  3.3588 0.0007828 ***
`Log(diff user) after` -0.21847715  0.03606764 -6.0574 1.383e-09 ***
`Market return after`   0.98222194  0.02265303 43.3594 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1.78
Residual Sum of Squares: 0.90724
R-Squared:      0.49031
Adj. R-Squared: 0.48981
Chisq: 1959.52 on 2 DF, p-value: < 2.22e-16
```

Does the number of Robinhood's subscribers influence the stock prices?

Does the number of Robinhood's subscribers influence the stock prices?

Annex P Output of F test after lockdown

```
            F test for individual effects

data:  `Stock Return after` ~ `Log(diff user) after` + `Market return after`
F = 2.2723, df1 = 19, df2 = 2018, p-value = 0.001338
alternative hypothesis: significant effects
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex Q Output of Breusch-Pagan Lagrange Multiplier test after lockdown

```
      Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels

data:  `Stock Return after` ~ `Log(diff user) after` + `Market return after`
chisq = 12.518, df = 1, p-value = 0.000403
alternative hypothesis: significant effects
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex R Output of Hausman test after lockdown

```
      Hausman Test

data:  `Stock Return after` ~ `Log(diff user) after` + `Market return after`
chisq = 6.5676, df = 2, p-value = 0.03749
alternative hypothesis: one model is inconsistent
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex S Output of the autocorrelation test after lockdown

```
      1) Durbin-Watson test for serial correlation in panel models

data:  `Stock Return after` ~ `Log(diff user) after` + `Market return after`
DW = 1.9938, p-value = 0.452
alternative hypothesis: serial correlation in idiosyncratic errors

      2) Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  `Stock Return after` ~ `Log(diff user) after` + `Market return after`
chisq = 0.016447, df = 1, p-value = 0.898
alternative hypothesis: serial correlation in idiosyncratic errors
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex T Output of the homoscedasticity test after lockdown

```
Breusch-Pagan test

data:  fixedeff.2
BP = 1052.5, df = 2, p-value < 2.2e-16
```

Does the number of Robinhood’s subscribers influence the stock prices?

Does the number of Robinhood's subscribers influence the stock prices?

Annex U Output for the fixed effects model after lockdown, after controlling for the heteroscedasticity in the model

```
t test of coefficients:

                        Estimate Std. Error t value Pr(>|t|)
`Log(diff user) after` -0.23254    0.12195 -1.9069  0.05667 .
`Market return after`   0.98164    0.03164 31.0257  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex V Output of unit root test before lockdown

1) Stock Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 2 lags (max: 5)
statistic (Wtbar): -107.92
p-value: 0
```

2) User Holding Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 0 - 3 lags (max: 5)
statistic (Wtbar): -58.402
p-value: 0
```

3) Market Return

```
Im-Pesaran-Shin Unit-Root Test
Exogenous variables: Individual Intercepts and Trend
Automatic selection of lags using SIC: 2 - 2 lags (max: 5)
statistic (Wtbar): -43.78
p-value: 0
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex W Output for the pooling model before lockdown

```
Pooling Model

Call:
plm(formula = `Stock Return before` ~ `Log(diff user) before` +
    `Market return before`, data = dados.p3, model = "pooling")

Balanced Panel: n = 20, T = 460, N = 9200

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-0.25090701 -0.00653584 -0.00047383  0.00588999  0.21867080

Coefficients:
                           Estimate  Std. Error  t-value  Pr(>|t|)
(Intercept)              0.00114818  0.00015733   7.2981 3.161e-13 ***
`Log(diff user) before` -0.50017548  0.01761211 -28.3995 < 2.2e-16 ***
`Market return before`   1.08657231  0.01072363 101.3250 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    4.5268
Residual Sum of Squares: 2.0414
R-Squared:      0.54905
Adj. R-Squared: 0.54895
F-statistic: 5598.85 on 2 and 9197 DF, p-value: < 2.22e-16
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex X Output for the fixed effects model before lockdown

```
Oneway (individual) effect Within Model

Call:
plm(formula = `Stock Return before` ~ `Log(diff user) before` +
    `Market return before`, data = dados.p3, model = "within")

Balanced Panel: n = 20, T = 460, N = 9200

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-0.25180197 -0.00642311 -0.00049893  0.00576242  0.21883699

Coefficients:
                         Estimate Std. Error t-value  Pr(>|t|)
`Log(diff user) before` -0.505742   0.017695 -28.581 < 2.2e-16 ***
`Market return before`   1.086502   0.010723 101.326 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    4.5245
Residual Sum of Squares: 2.0368
R-Squared:      0.54982
Adj. R-Squared: 0.54879
F-statistic: 5604.78 on 2 and 9178 DF, p-value: < 2.22e-16
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex Y Output for the random effects model before lockdown

```
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = `Stock Return before` ~ `Log(diff user) before` +
    `Market return before`, data = dados.p3, model = "random")

Balanced Panel: n = 20, T = 460, N = 9200

Effects:
                    var   std.dev share
idiosyncratic 0.0002219 0.0148972     1
individual    0.0000000 0.0000000     0
theta: 0

Residuals:
       Min.     1st Qu.      Median     3rd Qu.        Max.
-0.25090701 -0.00653584 -0.00047383  0.00588999  0.21867080

Coefficients:
                           Estimate  Std. Error  z-value  Pr(>|z|)
(Intercept)              0.00114818  0.00015733   7.2981 2.919e-13 ***
`Log(diff user) before` -0.50017548  0.01761211 -28.3995 < 2.2e-16 ***
`Market return before`   1.08657231  0.01072363 101.3250 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    4.5268
Residual Sum of Squares: 2.0414
R-Squared:      0.54905
Adj. R-Squared: 0.54895
Chisq: 11197.7 on 2 DF, p-value: < 2.22e-16
```

Does the number of Robinhood's subscribers influence the stock prices?

Does the number of Robinhood's subscribers influence the stock prices?

Annex Z Output of F test before lockdown

```
F test for individual effects

data:  `Stock Return before` ~ `Log(diff user) before` + `Market return before`
F = 1.071, df1 = 19, df2 = 9178, p-value = 0.3741
alternative hypothesis: significant effects
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex AA Output of Breusch-Pagan Lagrange Multiplier test before lockdown

```
      Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels

data:  `Stock Return before` ~ `Log(diff user) before` + `Market return before`
chisq = 0.0016089, df = 1, p-value = 0.968
alternative hypothesis: significant effects
```

Does the number of Robinhood’s subscribers influence the stock prices?

Annex AB Output of Hausman test before lockdown

```
      Hausman Test

data:  `Stock Return before` ~ `Log(diff user) before` + `Market return before`
chisq = 10.558, df = 2, p-value = 0.005098
alternative hypothesis: one model is inconsistent
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex AC Output of the autocorrelation test before lockdown

```
      1) Durbin-Watson test for serial correlation in panel models

data:  `Stock Return before` ~ `Log(diff user) before` + `Market return before`
DW = 2.0297, p-value = 0.9231
alternative hypothesis: serial correlation in idiosyncratic errors

      2) Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  `Stock Return before` ~ `Log(diff user) before` + `Market return before`
chisq = 2.7591, df = 1, p-value = 0.0967
alternative hypothesis: serial correlation in idiosyncratic errors
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex AD Output of the homoscedasticity test before lockdown

```
Breusch-Pagan test

data:  ols.3
BP = 2381.4, df = 2, p-value < 2.2e-16
```

Does the number of Robinhood's subscribers influence the stock prices?

Annex AE Output for the pooling model before lockdown, after controlling for the heteroscedasticity in the model

```
t test of coefficients:

                            Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)               0.00114818  0.00016989  6.7582 1.482e-11 ***
`Log(diff user) before` -0.50017548  0.07119336 -7.0256 2.283e-12 ***
`Market return before`   1.08657231  0.02415493 44.9835 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```