INSTITUTO
UNIVERSITÁRIO
DE LISBOA

The relationship between firm size and volatility of stock returns

Beatriz Franco Paralta da Silva Freire

Master in Finance

Supervisor:
Prof. Dr. José Dias Curto, Associate Professor with Aggregation,
ISCTE-IUL Business School

October, 2020

BUSINESS
SCHOOL

Quantitative methods for Management and Economics Department

The relationship between firm size and volatility of stock returns

Beatriz Franco Paralta da Silva Freire

Master in Finance

Supervisor:
Prof. Dr. José Dias Curto, Associate Professor with Aggregation,
ISCTE-IUL Business School

October, 2020

Acknowledgements

I would like to leave a special thank you to my supervisor José Dias Curto, for all the support and recommendations to improve this work. Without his help it would never have been possible to finish this process.

I also want to thank my mother and my grandmother for never letting me give up, always motivating me on the more challenging days, and sometimes believing in myself more than I do. Thank you for inspiring me and making me the person I am today. I will never have enough words to describe how important you are to me.

Last but not the least I want to thank my friends who have supported me throughout these years. Thank you for your patience and for trying to calm me down in the most complicated periods.

Abstract

The relationship between risk and returns its already very popular in financial markets, being the center of hundreds of studies nowadays. Within this wide range of researches, the firm size presents some impact on this relationship, being the smaller firms considered more risky and therefore, tending to reward investors with higher returns. But after all, to what extent is the firm size related with the volatility of such returns? Until today, few authors focused on this topic, motivating our study.

In representation of the small and large firms we considered the Russell 2000 and Russell 1000 indexes and concluded that the behavior of the returns differs between these two types of firms. According to our empirical evidence, the small firms are more volatile on the short-run, while the larger firms appear to be more affected by the shocks in a long-term perspective. Quite surprisingly, the negative shocks seem to affect more large firms than the smaller firms. Additionally, we focused our attention on the volatility spillovers across firms and, through the analysis of the FEVD, we concluded that the error variance of the Russell 1000 contributes to explain the error variance of the Russell 2000 in 84.78849%. Based on the DCC model, we confirmed the presence of volatility transmissions, since the conditional correlation tend to increase in periods of crisis.

Key words: Volatility, Volatility spillover, Firm size, Heteroscedasticity models

JEL classification: C58, G17

Resumo

A relação entre risco e retorno já é bastante popular nos mercados financeiros, sendo o centro de centenas de estudos hoje em dia. Dentro deste vasto leque de estudos, a dimensão das empresas apresenta implicâncias nesta relação, sendo que empresas mais pequenas são consideradas mais arriscadas e por isso tendem a recompensar os investidores com um nível de retorno mais elevado.

Mas afinal, até que ponto é que a dimensão da empresa está relacionada com a volatilidade dos retornos? Até agora são poucos os autores que se concentraram nesta relação temporal, motivando assim o nosso estudo.

Em representação das pequenas e grandes empresas considerámos os índices Russel 2000 e Russell 1000, e concluímos que o comportamento dos retornos difere entre tipos de empresas. De acordo com a nossa evidência, as pequenas empresas têm tendência a ser mais voláteis a curto prazo, enquanto que no longo-prazo as grandes empresas são mais afetadas pelos choques. Curiosamente, as más noticias parecem afetar mais as grandes empresas do que as empresas de menor dimensão. Adicionalmente, concentrámos a nossa atenção nas transmissões de volatilidade entre empresas e, através da análise da FEVD, concluímos que o erro da variância do Russell 1000 contribui para explicar em 84.78849% o erro da variância do Russell 2000. Com base no modelo DCC percebemos que existem transmissões de volatilidade, uma vez que as correlações condicionais tendem a aumentar em períodos de crise.

Palavras-chave: Volatilidade, Transmissão de volatilidade, Capitalização bolsista, Modelos de heterocedasticidade

Classificação JEL: C58, G17

Index

Index of Tables

Index of Figures

Glossary

ACF- Autocorrelation Function

ADF- Augmented Dickey-Fuller

AIC- Akaike Information Criterion

APARCH- Asymmetric Autoregressive Conditional Heteroskedasticity

AR- Autoregressive

ARCH- Autoregressive Conditional Heteroskedasticity

ARIMA- Autoregressive Integrated Moving Average

ARMA- Autoregressive Moving Average

BIC- Bayesian Information Criterion

CCC- Constant Conditional Correlation

DCC- Dynamic Conditional Correlation

EGARCH- Exponential Generalized Autoregressive Conditional Heteroskedasticity

FDDC- Flexible Dynamic Conditional Correlation

FEVD- Forecast Error Variance Decomposition

GARCH- Generalized Autoregressive Conditional Heteroskedasticity

GED- Generalized Error Distribution

GJR- Glosten-Jagannathan-Runkle

KPSS- Kwiatkowski–Phillips–Schmidt–Shin

MA- Moving Average

MAE- Mean Absolute Error

MAPE- Mean Absolute Percentage Error

MPE- Mean Percentage Error

MSE- Mean Square Error

PACF- Partial Autocorrelation Function

PP- Phillips-Perron

QQ- Quantile-Quantile

RMSE- Root Mean Square Error

SIC- Schwarz Information Criterion

SSR- Sum of Squares of the estimated Residuals

TGARCH- Threshold Autoregressive Conditional Heteroskedasticity

VAR- Vector Autoregressive

WN- White Noise

1 Introduction

The concern about the study of the stock market volatility arrives from its impact on investment, being crucial to understand how the market behaves in order to make better choices and manage risk (De Santis & İmrohoroğlu, 1997). The stock prices fluctuations reflect *"(...) changes in various aspects of our society such as economic, political, monetary, and so forth.*" (Gregoriou, 2009, p.5).

Nowadays the stock market is a subject of exhaustive research and there is an endless number of investigations dedicated to the study of the trade-off between expected return and risk of stock prices. Banz (1981) proposed the firm size as a driver of this relationship. Besides this purpose, the firm size was used for many others such as the study its impact on the relationship between both earnings and cash flows, and security returns (Charitou, Clubb, & Andreou, 2001), its impact on the exports (Bonaccorsi, 1992; Calof, 1994) and on organizational structure (Child, 1973; Pugh, Hickson, Hinings, & Turner, 1969), the study of its relationship with profits (Stekler, 1964; Samuels and Smyth, 1968), among others.

It is already known that the firm size can be seen as a measure of risk (Berk, 1995) but, besides the approach to this variable being very extensive, few have focused on relating it with the volatility of stock returns.

Based on the few studies available, the volatility of returns tends to vary across time and across different types of firms (Cheung & Ng, 1992). This leads us to the fundamental questions: to what extent is the market capitalization related to returns volatility? Are the returns of small firms more volatile than large firms? The main objective of this dissertation is to fill this lack in literature and be capable of answering these questions, along with the study of possible volatility spillovers between these two types of firms. More specifically, we aim to study the behavior of U.S. stock market and see if small stocks react differently from large stocks, and if these shocks are transmitted across firms.

By compiling the most important literature review conclusions, we defined all our study objectives. Therefore, we aim to confirm if:

- Small firms present more autocorrelation than large firms (Fisher, 1966);
- The shocks are more persistent on larger firms (Chelley-Steeley & Steeley, 1995);
- Impact tend to be higher to small firms (Chelley-Steeley & Steeley, 1995);
- Asymmetric impact of news is stronger for portfolios constituted by small firms (Chelley-Steeley & Steeley, 1996);

- Volatility spillover across firms: large firms volatility play an important role in predicting both volatility and mean return of smaller firms (Chelley-Steeley & Steeley, 1996);
- Negative relationship between the stock price volatility and firm size (Baskin, 1989; Habib, Kiani, & Khan, 2012; Hussainey, Mgbame, & Chijoke-Mgbame, 2011; Nazir, Nawaz, Anwar, & Ahmed, 2010).

To obtain these conclusions we will consider the GARCH models because, as we will see at the end of the literature review, due to the characteristics of stock returns, the standard deviation is not the most adequate to estimate stock returns volatility. To conclude about the volatility of returns of the different types of firms in study, we will consider the multivariate GARCH models.

In section 2 we have the literature review which starts by introducing the stock returns stylized facts, and by giving a background on the factors that may affect the stability of stock market volatility. Next, we study the behavior of the prices and understand the most popular reasons for such events to happen and confirm if there is evidence of an association to market capitalization (small and large firms). In the end, we see that in time series problems the normal standard deviation is not recommended and introduce the best methods to do the estimation of the volatility.

In section 3 we present the methodology where we describe the models mentioned previously, and the statistical tests we are going to use to obtain our conclusions. Finally, in section 4, we have the empirical study where we present the results and conclusions of our study.

2 Literature review

The financial asset prices and returns entail some well-known characteristics, mentioned as stylized facts, that are important to be aware of (Cont, 2001):

- No autocorrelation;
- Fat tails and Conditional Fat tails: the distribution of returns show pareto tails, before and after accounting the conditional variance models (less heavy after accounting for these models);
- Gain/loss asymmetry: there are more losses than gains;
- Aggregational Gaussianity: the distribution of a high frequency sample returns tend to proxy normality;
- Intermittency: the returns tend to exhibit high fluctuations;
- Volatility clustering: the absolute and squared returns show autocorrelation or, by other words, show some persistence in these nonlinear functions of returns. This is a sign of the existence of volatility clustering;
- Leverage effect: reflects the nonlinear dependence of returns and portraits a negative relationship between returns and the level of volatility;
- Autocorrelation of absolute returns decay slowly to zero;
- Volume/volatility relationship.

As seen by Mandelbrot (1963) and Fama (1965) the returns seem to fit better a stable Paretian distribution with exponent < 2, as firstly suggested by Mandelbrot (1963), than a Gaussian distribution. This means that a fat-tailed, or leptokurtic distribution, is more reasonable to describe the distribution of returns, as evidence proves that there is a larger number of observations in the tails of the distribution.

In addition to these characteristics, Cont (2001) also defends the importance verifying the presence of two more characteristics: stationarity and ergodicity. According with him, the statistical properties of the asset returns must not depend on time (stationarity), and the observations must be independent and identically distributed (ergodicity) to be adequate to forecast it. Enders (2015) calls a series stationary when the long-term mean and variance are constant, having no trend. In other words, stationarity pictures the long-term reversion to the mean and, according to the theory of efficient markets, this term is violated and we observe the presence of non-stationarity when the series has a unit root. This is a very common characteristic of the financial asset prices, unlike the returns, that tend to follow stationarity (Pagan, 1996). So, a non-stationary series can exhibit two different types of trend: deterministic trend (trend-stationary series) and stochastic trend (difference-stationary series). If the trend is deterministic the best method is to remove the trend by detrending, that is, "*regressing a variable on a constant and time and saving the residuals*". On the other hand, if the trend is stochastic, revealed by the presence of a unit root, we remove it by differencing the series (Enders, 2015, p.189).

The volatility clustering stylized fact is one of the most popular and it is the technical term to the reaction of the stock market to the shocks: "*... good news may tend to be followed more often by good news, and bad news may tend to be followed more often by bad news than by good news.*" (Fama, 1965, p.37). This behavior of returns is also known as ARCH effect, and its associated either to new information in the market, which tends to arrive in clusters, or to the reaction of the market to such information (Engle, Ito, & Lin, 1990).

According to the Efficient Market Hypothesis, changes in prices have a random-walk behavior, meaning they are not correlated with past values (Enders, 2015). This lack of autocorrelation of returns reflect the liquidity of the markets (Cont, 2001), as in efficient markets the successive price changes are independent (Fama, 1965).

The stock market has been exhaustively studied over the years and there are nowadays innumerable investigations available on volatility of financial markets, which discuss the reasons for the fluctuations of prices.

Considering a widespread data including the Civil War, World War I, the Great Depression, World War II, the OPEC oil shock, and the post-1979 period, Schwert (1988) observed that the stock market volatility recorded the higher levels of volatility during these periods of crisis.

One of the reasons studied by Schwert (1988) for the lack of stability in the stock market was Corporate Profitability, through the analysis of three variables: Payout Ratio, the Dividend Yield, and the Earnings Yield. Concerning the Payout Ratio, he observes that between 1929 and 1940 both Payout Ratio and stock volatility increased while, in the period 1973-1940, the variables had opposite behaviors. These distinct behaviors suggest that there is not a reliable trend to assure the effectiveness of a relation between the Payout Ratio and the stock market volatility. In terms of both Dividend and Earnings Yield, there was no evidence proving any trend at all. These results lead to the conclusion that the proof supporting that corporate profitability is related to market volatility is not very strong.

The next proposed reason regarded changes in Macroeconomic Variables (Inflation, Money Growth, Industrial Production, Bank Clearings, and Liabilities of Business Failures). Schwert (1988) found that, besides these variables being susceptible to higher levels of volatility in periods of crisis, only the Industrial Production contributes to understanding stock market volatility. In general, their evidence supports that the macroeconomic variables are not sustainable sources in the contribution to explain stock returns volatility. Considering stocks of New York Stock Exchange, Officer (1973) tries to relate the market factor (returns on stocks) to Business Fluctuations, the Formation of Securities and Exchange Commission, the Effect of Margin Requirements, and the increase in the number of stocks listed on the New York Stock Exchange. In accordance with Schwert (1988), he found that the only variable whose volatility helped explain the volatility in market factor was the Business Fluctuations, measured by Industrial Production.

In terms of stock market trading activity, periods of high trading activities correspond to periods with higher level of volatility (Schwert, 1988). This is in agreement with a study made by French & Roll (1986) who found that, during the period of January 1963 and December 1982, the stocks listed on the New York and American Exchanges recorded 70 times more volatility per hour on trading days than on non-trading days. The possibility of an association of trading and non-trading volatility to the firm size was studied but unfortunately, it was not found any relation. Amihud & Mendelson (1987) go further and study the level of volatility in two different trading mechanisms: opening transitions (also known as periodic clearing house) and closing transactions (also denoted by continuous dealership market). For 29 of 30 Dow Jones stocks, they conclude that, due to accumulated information, the variance of the opening transaction is 20% higher than closing transactions, existing, therefore, an higher level of volatility at the periodic clearing house.

This leads to the next argument: periods of high volatility are related to periods with an increase in information (Amihud & Mendelson, 1987; Beaver, 1968; French & Roll, 1986; Christie, 1982). Some events as earnings announcements are seen as moments that reveal enough information to be capable of influencing positively changes in prices and trading volume (Beaver, 1968). French & Roll (1986) discusses the higher volatility recorded on business days and defends that the Public Information, Private Information, and Pricing Errors are the main reasons for this evidence, as all these moments tend to happen during the business hours. Taylor (2005, p.22) states that there are several ways of news impacting prices: "*(...) relevant news about the asset and its cash flows, macroeconomic news, divergent beliefs about the interpretation of the news, and changes in investor sentiment.*".

One of the most studied relationships in finance is the association of the level of volatility to stock returns, and the majority of authors reach the same conclusion: stock returns and future volatility are negatively related (Black, 1976; Christie, 1982; Beckers, 1980; Cheung & Ng, 1992). The most popular reason to justify the association of periods with high volatility to periods where prices stand at a lower level is the leverage effect proposed by Black (1976). As the equity value is more sensitive to changes in the prices than the debt value, a decrease in the price will cause an increase in leverage effect, turning the stocks more risky (Beckers, 1980). The leverage effect states that bad and good news have different impacts on volatility, introducing an asymmetry of the news impact (Engle and Ng, 1991). Accordingly, Koutmos & Saidi (1995) affirm that bad news (associated to a decrease in prices) tend to affect the volatility 2.3 times more than good news. In conformity with these results, during recessions, the prices of stocks tend to decrease, influencing positively the leverage effect and increasing the volatility of stock returns (Schwert, 1988).

Besides the leverage effect, the contribution of the riskless interest rate to explain the changes in prices seemed relevant to Christie (1982). He found that the volatility of equity depends positively on the riskless interest rate or, in other words, the higher the riskless interest rate, the lower the leverage ratio and the higher will be the volatility of equity.

As the leverage effect is not enough to explain the negative relation between stock returns and future volatility ("predictive asymmetry"), the volatility feedback effect, also referred to as "no news is good news", aims to explain this feature along with the negative skewness ("contemporaneous asymmetry") and the excess kurtosis (Campbell & Hentschel, 1992). The volatility feedback states that all large news (good or bad) cause an increase in volatility, increasing the returns, and lowering the prices. The main difference between good and bad news is that, for negative shocks, this effect is more intense, generating excess kurtosis.

Unlike them, Poterba & Summers (1986) do not agree that volatility shocks have a higher impact on prices and defends that volatility is not persistent and that an increase of the market volatility by 50% only decreases the share prices by 11%, supporting that the market volatility is not efficient on explaining changes in prices.

The next possible explanation for the stock market volatility behavior is the time-varying risk premia considered by French, Schwert, & Stambaugh (1987). They concluded that the expected risk premium displayed a positive and negative relation with expected and unexpected changes in volatility, respectively.

Being the relationship between risk and return one of the most studied fields in finance, the inclusion of the firm size raised some interest to numerous authors.

The variables suggested to proxy the firm size, also referred as market value, are by now very extensive and there are a lot of different possible measures such as, for example:

1. Total Assets (Agarwal, 1979; Al-Khazali & Zoubi, 2005; Habib et al., 2012; Hopkins, 1988; Kimberly, 1976; Shalit & Sankar, 1977);
2. Total Sales (Agarwal, 1979; Hopkins, 1988; Kimberly, 1976; Shalit & Sankar, 1977);
3. Book Value of Equity (Al-Khazali & Zoubi, 2005; Shalit & Sankar, 1977);
4. Capacity (Kimberly, 1976);
5. Employment (Agarwal, 1979; Hopkins, 1988; Kimberly, 1976; Shalit & Sankar, 1977);
6. Number of Clients (Kimberly, 1976);
7. Market Capitalization (Al-Khazali & Zoubi, 2005; Berk, 1995; Vuolteenaho, 2002), defined by Hussainey, Mgbame, & Chijoke-Mgbame (2011) as the multiplication of the share price by the number of shares issued.

Kimberly (1976) found that more than 80% of the studies included in his research used the number of employees to measure firm size. Hopkins (1988) defends that, besides the firm size being an extensive concept, the firm assets and employment level are the most considered, while Shalit & Sankar (1977) argue that assets and stockholders equity are better measures for firm size rather than employment and sales. Hopkins (1988) states that, for all industries, the decisive mark to distinguish the size of the firm is assets value of $200,000,000 while for Fraser (1996), the differential mark stands at a market capitalization of £100,000,000. That is, firms with less than $200,000,000 in assets or less than £100,000,000 of market capitalization are characterized as small firms.

The interchangeability of these measures was extensively studied and the results differ substantially from author to author. As referred by Hopkins (1988, p.100) "*Interchangeability cannot be assumed a priori in all cases.*". Considering a data set of 6057 firms of ten different industries for the period from 1999 to 2002, Al-Khazali & Zoubi (2005) conclude that the adequate variable to define firm size depends on the business industry and that is why the conclusions about the best measure are very distinct. Additionally, Shalit & Sankar (1977) believe that the variables that adequately measure firm size depend on the purpose of the study.

According to Berk (1995), the firm size, the earnings to price ratio, the dividend yield, leverage, and the book-to-market equity are good variables to accurately explain future asset returns variation as they are cross-sectionally correlated with returns. Besides this, he also defends that the firm size can be seen as a measure of risk, as large firms tend to have larger market values, rising expected returns and lowering riskiness.

Between 1936 and 1975, stocks of small firms show higher returns as a way of compensating investors for the higher degree of uncertainty comparatively to large firms. This adjustment of returns to the level of risk to which the investors are exposed is entitled as size effect. Nevertheless, it is not proved that the market value of the firm is the reason for such relationship or if there is another reason linked to it (Banz, 1981). Besides the literature about this trade-off being extensive, these studies focus only on the expected returns and not on its volatility.

Although few, some researchers were interested in this question and have focused on studying the temporal relation between firm size and volatility returns. Some authors, not primarily worried about the relationship between firm size and returns volatility, found a negative impact of firm size on volatility.

The relationship between dividend policy and stock volatility was the focus of many studies and all authors used the same approach: consider size, debt, earnings volatility, asset growth and payout ratio as control variables (Baskin, 1989; Habib et al., 2012; Hussainey et al., 2011; Nazir et al., 2010). The size of firms was considered to be an important variable to include in this study because of its capacity of impacting price volatility through dividend policy. As larger firms are more diversified and, thereby, less affected by changes in individual markets while smaller firms have less information available, increasing the risk to which investors are exposed, the choice of dividends will differ across different types of firms. This introduces a firm size-price volatility relationship and justifies its inverse relation, defending the capacity of the firm size of controlling price volatility (Baskin, 1989). This evidence is supported for 2344 U.S. public corporations (Baskin, 1989), for the UK stock market (Hussainey et al., 2011), and for the Pakistan Stock Exchange (Habib et al., 2012; Nazir et al., 2010). Thus, according to these authors, the stock price volatility tends to increase with the decrease of firm size. Following Nazir, Nawaz, Anwar, & Ahmed (2010), the main reason for this is the lack of diversification of small firms, which causes the stocks to be less liquid.

Cheung & Ng (1992) were the first researchers mainly focused on investigating the relationship between firm size and returns volatility. Using an AR(1)-EGARCH(1,2)-in-mean model, they observed that negative impacts on prices have distinctive effects on firms with different market capitalizations. These negative shocks carry more dramatic consequences for small capitalization firms, triggering an higher level of volatility. In order to analyze the relationship between two variables (being, in this study, the firm size, the debt-to-equity ratio, and the coefficients of the model), they consider the nonparametric test Spearman rank correlation coefficients. Based on this test, they prove the existence of a negative relation between debt-to-equity ratio and firm size, and a negative relation between D/E ratio and θ_i (coefficient that measures the relationship between volatility and prices). Shortly, they support that negative shocks have more impact than positive shocks, and that smaller firms do not have the same capability to face these shocks, being more volatile than the larger firms.

Although the GARCH models are the most appropriate to this type of problem, Duffee (1995) chose to use the normal standard deviation and then, applied the Spearman Rank Correlation Coefficient to measure the level of correlation of the firm size, debt-to-equity ratio, and the firm volatility. Based on this coefficient, the results lead to the same already described: negative relation between firm size and volatility, and positive relation between debt-to-equity ratio and volatility. His main conclusions were that the lower the firm, the higher the debt-to-equity ratio and the higher will be the volatility of the firm. In accordance with Black (1976), Christie (1982) and Cheung & Ng (1992), they found a negative relationship between the stock returns and future changes in stock returns volatility, and positive relationship between stock returns and today's volatility, revealing positive skewness in the returns. Duffee (1995) experienced that both of these conclusions concerning the relationship between stock returns and changes in volatility were stronger for firms with lower market capitalization. Even though he agrees with the most popular explanation for the negative relationship presented before, he is not sure the leverage effect is the only reason for such event, ending up emphasizing that there must be other factors that associate the debt-to-equity ratio to the stock returns-future changes in stock returns volatility relationship.

Indexes or portfolios of small firms tend to present more serial correlated returns than large firms because, as concluded by Fisher (1966), these types of firms are less active in trading than larger firms. So, comparing with large firms, the prices of small firms tend to adjust slower to new information, revealing non-synchronous price adjustments. This lack of studies including the returns autocorrelation structure encouraged Chelley-Steeley & Steeley (1995, 1996) to study the conditional mean and variance of returns. Their main goal was to analyze four portfolios constituted by UK stocks sorted by market capitalization, between January 1976 and December 1991. In order to face autocorrelation, and to see if there was any association between the impact and persistence of shocks and the firm size, they used an ARMA(1,1)-GARCH(1,1)-M model to estimate the conditional mean and variance of returns. The role of information on current volatility appears to be more critical to portfolios constituted by smaller firms than large ones. As the former are firms that do not trade as frequently, they seem to have more leptokurtic and autocorrelated returns and be more affected by shocks on volatility (Chelley-Steeley & Steeley, 1996) . Thus, the impact tend to be higher to small firms, as stated by Chelley-Steeley & Steeley (1995, p.435) "*(...) the volatility of small firm portfolio returns is more clustered than the volatility of large firm portfolio returns*.", but more persistent for large firm portfolios. As confirmed by the analysis of the half-life of the shocks, the duration is only seven trading days for the small firm portfolios while, for large firm portfolios, it is more than the double, reaching twenty trading days. This increased duration is associated with mechanical rules in trading.

As already referred by previous authors, the leverage effect also seemed to change under market capitalization differences. According to Chelley-Steeley & Steeley (1996), this asymmetric impact of news is stronger for portfolios constituted by small firms.

Additional to these findings, Chelley-Steeley & Steeley (1996) also experienced events of volatility spillover across firms. In line with a study made by Conrad, Gultekin, & Kaul (1991), large firms volatility play an important role in predicting both volatility and mean return of smaller firms.

Volatility spillovers are also mentioned by Harris & Pisedtasalasai (2006) as transmissions of volatility between markets, and are considered useful events to conclude about market efficiency, portfolio management, and financial analysis (pricing). According with their study, it was recorded volatility spillovers between large and small firms, and vice versa. In the first case, the large firm's volatility affect positively the small firms. By contrary, the volatility effects of small firm's gave negative consequences on large firm's volatility. This means that an increase of small firms' volatility will imply a decrease in large firms' volatility. The main doubt stood at the reasons for such transmissions, remaining the uncertainty of the origins of these spillovers until Harris & Pisedtasalasai (2006) prove that what causes it is not the non-synchronous trading but the new information arriving to the market.

The main tools used to estimate and forecast assets volatility are the univariate models, whereas the first and most known are the Autoregressive Conditional Heteroskedasticity and Generalized Autoregressive Conditional Heteroskedasticity models. These models assume the lack of stability on variance that is, existence of heteroscedasticity, filling the gap of the traditional least squares model. Besides this, they also account for the stylized fact volatility clustering. So, we can say that these models emerged to meet the need of financial analysis, being used for innumerous ends: risk management, analysis and selection of portfolio or assets, assets pricing, among others (Engle, 2001). The Autoregressive Conditional Heteroskedasticity model of Engle (1982) emerged with the main objective of correcting a lack in modeling time variation and include higher order moments in the estimation of time-varying volatility. This model deals with volatility clustering and considers that the estimation of variance is a linear function of the past squared value. (Bollerslev *et. al*, 1992). Despite this model being revolutionary, Bollerslev (1986) quickly found a crucial disadvantage and decided to estimate the variance by giving less relevance to old information. Although these models have changed the world of volatility estimation, there was one important feature that was not considered: the asymmetric impact of news. To face this, there were developed innumerous of variants for the univariate models: Threshold Autoregressive Conditional Heteroskedasticity model of Zakoian (1994) , GJR model of Glosten, Jagannathan, & Runkle (1993), Asymmetric Power Autoregressive Conditional Heteroskedasticity model by Ding, Granger, & Engle (1993), among others.

Even after countless variants of the univariate class of models, there are some conclusions that can only be achieved by considering multivariate models. These models are considered to applications as asset pricing, portfolio management, hedging strategies, and Value-at-Risk forecast, and allow us to do a reliable analysis of the impact of some assets on the domestic and foreign assets, or market in general (Orskaug, 2009). Examples of these models are the Constant Conditional Correlation of Bollerslev (1990) and the Dynamic Conditional Correlation of Engle (2002).

3 Methodology

We will follow an approach similar to Chelley-Steeley & Steeley (1995, 1996) with the difference that, instead of creating two market-based portfolios, we will consider the adjusted-close prices for the past ten years of two indexes of Russel US index: Russell 2000 and Russell 1000.

A common way to treat the data is to divide it into two phases: in-sample and out-of-sample periods. This is helpful because a good in-sample model can be poorer and produce unsatisfactory out-of-sample results. So, the in-sample period is used to estimate the parameters of the model while the motivation behind the out-of-sample is to evaluate the performance of the forecast and see if the model produces reliable predictions. The sample size assigned to the out-of-sample part depends on the time horizon we want to predict, being the most common 20% of the data (Hyndman & Athanasopoulos, 2018).

3.1 Conditional Mean Models

"*To make a forecast is to infer the probability distribution of a future observation from the population, given a sample of past values.*". This has led to the need for tools capable of adequately estimate different stochastic processes. This lack of models was addressed by the stationarity stochastic processes (Autoregressive, Moving Average, Autoregressive Moving Average), and non-stationary processes (Autoregressive Integrated Moving Average) (Box *et al.*, 2016, p.19).

The Autoregressive (AR) Model, also referred to as autoregressive process of order p AR(p), consists of modeling a process $\tilde{z}_t$ based on the past values and a random innovation α_t:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + \alpha_t \tag{3.1}$$

$$\tilde{z}_t = z_t - \mu \tag{3.2}$$

The Moving Average (MA) Model, or moving average process of order q MA(q), defines the process $\tilde{z}_t$ as a finite function of the past α_t. The moving average process is given by:

$$\tilde{z}_t = \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \cdots - \theta_q \alpha_{t-q} \tag{3.3}$$

Combining these two components, we have the Autoregressive Moving Average (ARMA) process, a set of autoregressive and moving average terms:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \cdots + \phi_p \tilde{z}_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \cdots - \theta_q \alpha_{t-q} \quad (3.4)$$

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\tilde{z}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\alpha_t \quad (3.5)$$

$$\phi(B)\tilde{z}_t = \theta(B)\alpha_t \quad (3.6)$$

Regarding the stationarity and invertibility conditions of the ARMA model, it depends on the ones from the AR and MA models. In order to the AR process be stationary, the roots of the autoregressive operator $\phi(B) = 0$ must lie outside the unit circle (that is, their absolute value must be higher than 1). In terms of invertibility, to the MA process be invertible, the roots of $\theta(B) = 0$ must lie outside the unit circle. Following these, the ARMA process is stationary and invertible if the roots of $\phi(B) = 0$ and $\theta(B) = 0$ lie outside the unit circle, respectively.

However, the process can be non-stationary, leading to the need of a class of models that can estimate this type of data. The Autoregressive Integrated Moving Average (ARIMA) Model emerged as a non-stationary process, designed to deal with this very common characteristic of the financial time series data. This model introduces the integrated term d, which corresponds to the number of differences needed to transform the series into a series with stationary nature (in most cases, the first difference or, at maximum, the second difference, is enough to remove the trend (Box *et. al*, 2016)). The generalized autoregressive operator $\varphi(B)$ is a nonstationary autoregressive operator and can be written as a function of the stationary autoregressive operator $\phi(B)$:

$$\varphi(B) = \phi(B)(1 - B)^d \quad (3.7)$$

$$\varphi(B)z_t = \phi(B)(1 - B)^d z_t = \theta(B)\alpha_t \quad (3.8)$$

$$\phi(B)w_t = \theta(B)\alpha_t \quad (3.9)$$

Where the integrated part is represented by w_t, which is described as $w_t = (1 - B)^d z_t = \nabla^d z_t$.

So, the ARIMA (p,d,q) process, constituted by autoregressive, integrated and moving average terms, is given by:

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \cdots - \theta_q \alpha_{t-q} \quad (3.10)$$

3.2 Univariate Conditional Heteroskedastic Models

3.2.1 Autoregressive Conditional Heteroskedasticity (ARCH)

The Autoregressive Conditional Heteroskedasticity (ARCH) model of Engle (1982) is the pioneer of a wide range of models that considers that past values are rich in information useful to forecast conditional variance. This model is very respected in econometrics due to their predictive power (as it uses the past values to forecast the future), due to its capacity of estimating the volatility of an asset (revealing its usefulness in portfolio management), and its capability of capturing the ARCH effect caused by misspecification (due to omitted variables or structural changes) (Engle, 1982). It incorporates the well-known stylized fact volatility clustering and can be denoted by ARCH (q), where the q corresponds to the persistence of volatility (Bera & Higgins, 1993).

Engle (1982) defines the more general form of an ARCH process where the conditional volatility changes with past errors as:

$$y_t = \epsilon_t h_t^{1/2} \tag{3.11}$$

Assuming ϵ_t is independent and identically distributed (i.i.d) with mean $E(\epsilon_t) = 0$ and variance $V(\epsilon_t) = 1$

$$y_t | \psi_{t-1} \sim N(0, h_t) \tag{3.12}$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 \tag{3.13}$$

$$h_t = h(y_{t-1}, y_{t-2}, \dots, y_{t-p}, \alpha) \tag{3.14}$$

Where,

h_t: variance σ_t^2;

ψ_{t-1}: information at time t-1;

p: order of the ARCH process;

α: a vector of unknown parameters.

Assuming that the conditional mean of y_t corresponds to $x_t\beta$ that is, a linear combination of lagged endogenous and exogenous variables included in the information set ψ_{t-1} with β:

$$y_t | \psi_{t-1} \sim N(x_t\beta, h_t) \tag{3.15}$$

$$h_t = h(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-p}, \alpha) \tag{3.16}$$

$$\epsilon_t = y_t - x_t\beta \quad \text{where } \beta \text{ is a vector of unknown parameters} \tag{3.17}$$

As the information set can also consider both current and lagged x's, we can write the variance in a broader form:

$$h_t = h\left(\epsilon_{t-1}, \dots, \epsilon_{t-p}, x_t, x_{t-1}, \dots, x_{t-p}, \alpha\right) = h(\psi_{t-1}, \alpha) \qquad (3.18)$$

The coefficients α_i must be constant and nonnegative for all values of i. Regarding the stationarity of the series, the theorem 2 presented in the seminal paper of Engle (1982) states that for the variance of a first-order linear ARCH process be finite, that is, covariance stationary, the roots must lie outside the unit circle.

3.2.2 Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

Similar to the transition from AR to ARMA model, we have the evolution from ARCH model to Generalized Autoregressive Conditional Heteroskedastic (GARCH) model. The ARCH model requires a long and fixed lag structure in conditional variance, which could end up calling into question the validity of the nonnegativity constraints. So, the GARCH model emerged with the need of facing this undesirable characteristic and allowing the lag structure to be more flexible and have higher memory (Bollerslev, 1986). In other words, the GARCH is a more parsimonious model and represents an high or infinite order ARCH process (Bera and Higgins, 1993). According to Bollerslev (1986, p.309) "*In the ARCH* (q) *process the conditional variance is specified as a linear function of past sample variances only, whereas the GARCH* (p, q) *process allows lagged conditional variances to enter as well.*". The values of p and q determine the type of process we have, being transformed into other processes if some values are assumed: (1) If $p = 0$: the model corresponds to a simple ARCH (q) and, (2) if $p = q = 0$: the ε_t is white noise (WN). The general GARCH (p, q) is defined as:

$$\varepsilon_t | \psi_{t-1} \sim N(0, h_t) \qquad (3.19)$$

$$\varepsilon_t = y_t - x_t' b \qquad (3.20)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} = \alpha_0 + A(L)\varepsilon_t^2 + B(L) h_t \qquad (3.21)$$

Where,

ε_t: a real-valued discrete-time stochastic process;

ψ_{t-1}: information set at time t-1;

x_t': vector of explanatory variables;

b: vector of unknown parameters.

In order to the term ε_t be covariance stationary, the sum of the alpha and beta terms must be lower than 1 (Bollerslev, Chou, & Kroner, 1992). A common event in high frequency data is that this sum results in a value close to one, being an indicator that we have an integrated model and the volatility is highly persistent (Bollerslev, Engle, & Nelson, 1994).

In order to assure the validity of de model the coefficients must assume nonnegative values, demanding the following inequality restrictions:

$$\begin{gathered} p \geq 0, \qquad q > 0 \\ \alpha_0 > 0, \qquad \alpha_i \geq 0, \qquad i = 1, \dots, q \\ \beta_i \geq 0, \qquad i = 1, \dots, p \end{gathered} \tag{3.22}$$

However, some events as misspecification or sampling errors can drive to the negativity of coefficients, leading Nelson & Cao (1992) to investigate the flexibility of these inequality conditions. According to their study, these assumptions do not need to be so strict as demanded by Bollerslev (1986), and there are some cases where the nonnegativity of the conditional variance is not disputed:

- $p = 1$: According to the theorem 1 presented in their paper, the inequality constraints are only more flexible for α_i with $i \geq 2$. The remaining conditions must be fulfilled otherwise, the conditional variance will not assume positive values;
- $p = 2$: As in the previous case, the restrictions are more flexible and allow the β_1 to assume negative values;
- $p \geq 3$: In this case, the evidence is not so clear and is more challenging to prove.

3.2.3 Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH)

Engle (1982) tested an exponential form of the ARCH model and showed that the variance remained positive no matter which value is assumed by the coefficient alpha, unless when α_1 is different from 0, leading to a model with infinite variance. Given these results, he recognizes that this form of the model should be studied more carefully.

Besides the ARCH and GARCH models successfully accounting for volatility clustering, Nelson (1991) highlighted a few drawbacks which need to be considered in order to find a more reliable model:

- The inverse relationship between stock returns and volatility found by Black (1976) (in both ARCH and GARCH models the only important feature is the size and not the sign of the unanticipated stock returns);
- The coefficients of the model must assume values equal to or above 0;
- Hard to interpret the persistence of shocks to conditional variance.

Therefore, the Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) model hypothesized by Nelson (1991) arises as model capable of dealing with these three drawbacks. Instead of requiring nonnegative coefficients, the sign of the conditional variance σ_t^2 is approached differently and, in order not to compromise the sign of σ_t^2 and keeping it positive, the variance is modeled by taking the logarithm of σ_t^2. By doing this, the β_k can assume any value without questioning the sign of σ_t^2. The EGARCH can be represented by an infinite MA (33) or by an ARMA process (34):

$$ln\,(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k\, g(z_{t-k}) \qquad \beta_1 = 1 \tag{3.23}$$

$$ln\,(\sigma_t^2) = \alpha_t + \frac{(1 + \psi_1 L + \cdots + \psi_q L^q)}{(1 - \Delta_1 L + \cdots + \Delta_p L^p)}\, g(z_{t-1}) \tag{3.24}$$

Where $\{\alpha_t\}_{t=-\infty,\infty}$ and $\{\beta_k\}_{k=1,\infty}$ are real, nonstochastic, scalar sequences.

The asymmetric effect is measured by $g(z_t)$, which consists of the set of size and sign of z_t. In order to have accurate results, the $g(z_t)$ must be a linear combination of z_t and $|z_t|$:

$$g(z_t) = \theta z_t + \gamma[|z_t| - E|z_t|] \tag{3.25}$$

Where $\{g(z_t)\}_{t=-\infty,\infty}$ is a zero-mean, i.i.d random sequence and both θz_t and $\gamma[|z_t| - E|z_t|]$ have mean zero.

More specifically, the coefficient θ measures the asymmetric impact of shocks on changes on volatility. The asymmetric effect is granted by $g(z_t)$ because the slope depends on the value of z_t: if (1) $0 < z_t < \infty$, the function $g(z_t)$ is linear with slope $\theta + \gamma$, and if (2) $-\infty < z_t \leq 0$, the function $g(z_t)$ is linear with slope $\theta - \gamma$.

The segment $\gamma[|z_t| - E|z_t|]$ of $g(z_t)$ is responsible for measuring the magnitude/size effect:

- $\gamma > 0$ and $\theta = 0$: When the magnitude of z_t is above the expected value, there is a positive innovation in $g(z_t)$;
- $\gamma = 0$ and $\theta < 0$: The innovations in conditional variance and in returns have opposite signs.

Regarding z_t, it can assume three different distributions: normal distribution, student-t distribution, and generalized error distribution (GED). The density function of GED is described as:

$$f(z) = \frac{v\, exp\left[-\left(\frac{1}{2}\right)|z/\lambda|^{v}\right]}{\lambda 2^{(1+1/v)}\Gamma(1/v)} \tag{3.26}$$

Where Γ(.) is a gamma function, and the following conditions are necessary:

$$\begin{gathered} -\infty < z < \infty \\ 0 < v \leq \infty \end{gathered} \tag{3.27}$$

$$\lambda = \left[2^{(-2/v)}\Gamma(1/v)/\Gamma(3/v)\right]^{1/2} \tag{3.28}$$

The parameter v measures the type of the distribution of both z_t and errors:

- $v = 2$: The distribution of z_t is normal;
- $v < 2$: The distribution of z_t is leptokurtic;
- $v > 2$: The distribution of z_t is platykurtic.

The stationary and ergodicity of the EGARCH model depend on the persistence of shocks. In the GARCH model, the measurement of the persistence of shocks in variance is not clear and depends on different interpretations. In the case of the EGARCH, the $ln\ (\sigma_t^2)$ is said to be strictly stationary and ergodic if the shocks do not persist and if we eliminate the deterministic trend. The conditions for the strict stationarity and ergodicity are described on theorem 2.1 presented by Nelson (1991) which shortly states that, if $ln\ (\sigma_t^2)$ is described by a first-order autoregressive process with coefficient Δ, it is strictly stationary and ergodic if $|\Delta| < 1$.

3.2.4 GJR (Glosten- Jagannathan-Runkle)

Besides the studies about the risk and return tradeoff being extensive, the investors reward for the increased level of risk has not received so much attention temporally. Even so, the time behavior of this tradeoff was studied by a few researchers and the conclusions about the relationship between the expected returns and the conditional variance differ from author to author. Glosten, Jagannathan, & Runkle (1993) suspect that this disagreement is driven by the difference in methodology. The main distinction between the different studies available were the methods used whereupon, for the studies which was observed a positive or inexistent relationship, the GARCH-M model was used. Therefore, Glosten, Jagannathan, & Runkle (1993) recognize the GARCH-M model as an insufficient to model the data and generated the Glosten Jagannathan Runkle GARCH (GJR-GARCH) model, a modified version of the GARCH-M model. This version considers the EGARCH model of Nelson (1991) as the basis, complementing it with the following features:

- Deterministic seasonal dummies: January and October are considered months where the volatility tends to be higher. January is a more volatile month because 2/3 of the firms use the calendar year as a fiscal year, leading to an increase of information in beginning of the year. Regarding the month of October, it was included due to the October 1987 crash, but there is no plausible reason for this yet;
- Asymmetric effect on conditional variance;
- Nominal interest rate: The short-term nominal interest rate is considered to be a good auxiliary tool to the estimation of future volatility of excess returns as it can reflect the expectations of inflation.

According to Bollerslev, Engle, & Nelson (1994, p.2970), the GJR-GARCH model "*(...) allows a quadratic response of volatility to news with different coefficients for good and bad news, but maintains the assertion that the minimum volatility will result when there is no news.*":

$$\sigma_t^2 = w + \sum_{i=1,q} [\alpha_i^+ I(\varepsilon_{t-i} > 0)|\varepsilon_{t-i}|^2 + \alpha_i^- I(\varepsilon_{t-i} \leq 0)|\varepsilon_{t-i}|^2] + \sum_{j=1,p} \beta_j \sigma_{t-j}^2 \qquad (3.29)$$

Where $I(.)$ is an indicator function.

Additionally, regarding the properties of data, Glosten, Jagannathan, & Runkle (1993) take two more conclusions:

- Differing only on the frequency of the data, they disagree with Nelson (1991) and defend that the conditional volatility does not record much persistence. That is, the monthly data point out to the almost inexistence of persistence in volatility, in contrary to daily returns of Nelson (1991) that reveal a positive relationship between the frequency of data and persistence;
- Positive and negative unexpected returns have different impacts (positive returns lead to a decrease in volatility while negative returns lead to an increase in volatility) while, according to Nelson (1991), both positive and negative have the same consequences (higher volatility).

3.2.5 Asymmetric Power Autoregressive Conditional Heteroskedasticity (APARCH)

The main focus of Ding, Granger, & Engle (1993) was to analyze the autocorrelation of the series. They investigated the level of autocorrelation for different forms of the returns and, besides proving that the absolute returns were more autocorrelated than the squared returns, they did not found a conceivable motive for considering that the conditional volatility should be a linear function of the lagged squared residuals or the lagged absolute residuals.

By analyzing the power transformation $|r_t|^d$, with d assuming a broad range of values, they found significant positive autocorrelation at least up to lag 100, proving that the market stock returns have long-term memory. So, Ding, Granger, & Engle (1993) introduced a new and broader model that incorporates seven different models and a power transformation capable to deal with the long-term memory along with the leverage effect and volatility clustering already considered previously. More details about these models are presented in the appendix A of their paper. The new model was entitled Asymmetric Power Autoregressive Conditional Heteroskedasticity (APARCH) and is given by:

$$\varepsilon_t = \sigma_t e_t \qquad e_t \sim N(0,1) \tag{3.30}$$

$$\sigma_t^{\delta} = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^{\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-i}^{\delta} \tag{3.31}$$

Where the following conditions are demanded:

$$\begin{gathered}\alpha_0 > 0 \\ \alpha_i \geq 0 \; for \; i = 1, \dots, p \\ -1 < \gamma_i < 1 \; for \; i = 1, \dots, p \\ \beta_j \geq 0 \; for \; j = 1, \dots, q\end{gathered} \tag{3.32}$$

3.2.6 Threshold Autoregressive Conditional Heteroskedasticity (TGARCH)

The Threshold GARCH (TGARCH) model of Zakoian (1994) is very similar to the GJR model, allowing the σ_t to behave differently to both sign (the asymmetric response to shocks) and magnitude of the innovations. The main differentiator of this model is that it approaches the conditional standard deviation instead of the conditional variance and, therefore, can be described as:

$$\varepsilon_t = \sigma_t Z_t \tag{3.33}$$

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^+ \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^- + \sum_{j=1}^{p} \beta_j \sigma_{t-j} \tag{3.34}$$

Where,

Z_t is i.i.d and independent of $\underline{\varepsilon_{t-1}}$ for all t, with mean 0 and variance 1;

$\underline{\varepsilon_{t-1}}$: information at time t-1;

$(\alpha_i^+)_{i=1,q}$, $(\alpha_i^-)_{i=1,q}$ and $(\beta_j)_{j=1,p}$ are real scalar sequences;

The positive and negative parts of the real-value discrete-time process ε_t are described by $\varepsilon_t^+ = \max(\varepsilon_t, 0)$ and $\varepsilon_t^- = \min(\varepsilon_t, 0)$.

Considering that $\alpha_i^+ = \alpha_i^- = \alpha_i$, we can also write a new version of the TGARCH model. This adaptation is similar to the GARCH and it entails the same drawback, as the model developed by Bollerslev (1986), and only accounts for the magnitude of the shocks:

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i |\varepsilon_{t-i}| + \sum_{j=1}^{p} \beta_j \sigma_{t-j} \tag{3.35}$$

Zakoian (1994) estimated four different models (GARCH, threshold GARCH, exponential GARCH, and GARCH with absolute values) and concluded that both EGARCH and

TGARCH, the ones which similarly approach the asymmetric effect, achieve identical results. At the end, the TGARCH stands out due to the inclusion of the standard deviations instead of variances, which allow it to react more efficiently to large shocks. The main differences between the EGARCH model and the TGARCH are that in the latter the volatility is a function of nonnormalized shocks and the asymmetry effect can be different for the different lags (for example, $\alpha_1^+ - \alpha_1^- > 0$ while $\alpha_2^+ - \alpha_2^- < 0$).

3.3 Multivariate Conditional Heteroskedastic Models

3.3.1 Conditional Mean: Vector Autoregressive (VAR)

The Vector Autoregressive (VAR) model is a multivariate model composed by more than one time series, being more advantageous than the univariate models when we want to analyze the relationship between two or more series. We should take this model into account when we want to analyze the impact of one variable to another and see how much it contributes to its forecast (Granger causality test), do an impulse response analysis (see if a shock in one variable affects the other variable), or do a forecast error variance decomposition (see how much of the forecast variance of a variable affects another variable) (Hyndman & Athanasopoulos, 2018).

Enders (2015) defends that the VAR model is recommended when are there are doubts about the nature of the variables, being defined with the goal of including exogenous variables. If we consider a first-order VAR constituted by two stationary time series y_t and z_t, we can write our model as:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt} \tag{3.36}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt} \tag{3.37}$$

Where ε_{yt} and ε_{zt} must be white-noise and have no correlation. Regarding the impacts of series on the other, it is measured by the $-b_{12}$ and γ_{12} (for the impact of a change in z_t), and$-b_{21}$ and γ_{21} (for the impact of a change in y_t).

3.3.2 Constant Conditional Correlation (CCC)

The Constant Conditional Correlation (CCC) of Bollerslev (1990) is a multivariate heteroscedasticity model that approaches the conditional variance, conditional covariance, and conditional correlation. The CCC model considers that the conditional variance and covariance are not constant over time, but the conditional correlation is. This model proved to be a success in asset pricing, hedging strategies, and in other vast number of applications, with the main advantage being the easiness of its estimation. We can describe it by considering the Nx1 time-series vector y_t:

$$y_t = E(y_t|\psi_{t-1}) + \varepsilon_t \tag{3.38}$$

$$var(\varepsilon_t|\psi_{t-1}) = H_t \tag{3.39}$$

Where,

ψ_{t-1}: σ at time t-1;

ε_t: error term;

H_t: time-varying conditional covariance.

The conditional correlation between two assets can vary between -1 and 1 and is given by ρ_{ijt}:

$$\rho_{ijt} = \frac{h_{ijt}}{\sqrt{(h_{iit}h_{jjt})}} \tag{3.40}$$

Where h_{ijt} is the ij^{th} element of the matrix H_t.

If we rewrite this in order of the conditional variances h_{ijt}, and respect the conditions $j = 1, \dots, N$ and $i = j + 1, \dots, N$, we obtain the constant conditional correlations:

$$h_{ijt} = \rho_{ijt}\sqrt{(h_{iit}h_{jjt})} \tag{3.41}$$

Besides the definition on (3.41), the conditional covariance matrix can also be defined as:

$$H_t = D_t \Gamma D_t \tag{3.42}$$

Where D_t represents the matrix diagonal which, in turn, represents the variance of the assets σ_{1t}, …, σ_{Nt}, and Γ represents a NxN time invariant matrix with the elements $\rho_{ij}\sqrt{(\omega_i\omega_j)}$.

3.3.3 Dynamic Conditional Correlation (DCC)

The Dynamic Conditional Correlation (DCC) model surged with the requirement of models capable of capturing the time-varying propriety of correlations, as usually an increase in volatility can have an impact on the level of conditional correlation too (Billio, Caporin, & Gobbo, 2006). Therefore, DCC is very similar to the CCC, adapting only the constant conditional correlations rule. The main benefit of this model is that we can include as many parameters as we want, without relying on the number of series we have in our model (Engle, 2002). The approach is somewhat different from other multivariate models: the two main phases to estimate this model are (1) to estimate a univariate GARCH model, and (2) use its standardized residuals and proceed with the multivariate estimation. The definition of this model is very similar to the CCC, with the difference that here, the correlation matrix R_t, is not constant over time (Engle & Sheppard, 2001):

$$r_t|\mathcal{F}_{t-1} \sim N(0, H_t) \tag{3.43}$$

$$H_t = D_t R_t D_t \tag{3.44}$$

Where D_t represents the diagonal of the k x k matrix, which in turn represents standard deviations.

However, the standard DCC model did not consider the asymmetries of news, leading Cappiello, Engle, & Sheppard (2006) to originate the Asymmetric Generalized Dynamic Conditional Correlation (AG-DCC) model:

$$Q_t = (\bar{Q} - A'\bar{Q}A - B'\bar{Q}B - G'\bar{N}G) + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'Q_{t-1}B + G'n_{t-1}n'_{t-1}G \tag{3.45}$$

Where,

A, B, and G: matrix diagonal parameters;

n_t: $I[\varepsilon_t < 0] \circ \varepsilon_t$ with $\circ$ illustrating the Hadamard product;

$\bar{N}$: $I[n_t n'_t]$;

$\bar{Q}$: $T^{-1} \sum_{t=1}^{T} \varepsilon_t \varepsilon'_t$;

$\bar{N}$: $T^{-1} \sum_{t=1}^{T} n_t n'_t$.

In order to permit different correlations across assets, Billio, Caporin, & Gobbo (2006) also provided a new form of the DCC model called Flexible Dynamic Conditional Correlation (FDCC) model. The main drawback of this model is that the unconditional correlation is not considered. The conditional covariance matrix H_t is defined similarly, differing on the correlation matrix R_t representation:

$$H_t = D_t R_t D_t \tag{3.46}$$

$$R_t = (Q_t^*)^{-1} Q_t (Q_t^*)^{-1} \tag{3.47}$$

$$Q_t = cc' + aa' \circ \eta_t \eta_t' + bb' \circ Q_{t-1} \tag{3.48}$$

Where,

D_t: diagonal of the conditional standard deviations matrix, that is, $D_t = diag(\sigma_{11,t}, \sigma_{22,t}, \dots, \sigma_{kk,t})$;

Q_t^*: diagonal of the conditional correlation matrix, that is, $Q_t^* = diag(\sqrt{q_{11,t}}, \sqrt{q_{22,t}}, \dots, \sqrt{q_{kk,t}})$;

η_t: standardized residuals;

c, a, and b: k-dimensional vectors.

3.4 Statistical tests

3.4.1 Unit Root Tests

3.4.1.1 Augmented Dickey-Fuller Test

One of the most popular unit root tests is the Augmented Dickey-Fuller (ADF), an improved version of the Dickey-Fuller test of Dickey & Fuller (1979), which considers that the errors are independent and homoscedastic. The ADF test is a more general test that, contrary to the simple Dickey-Fuller test that is only suitable to the first-order autoregressive process, includes lagged changes and, therefore, is destined to higher-order autoregressive processes (Enders, 2015). Despite the usefulness of this test, it entails a few drawbacks highlighted by Enders (2015):

- The process needs to be correctly specified, including all autoregressive terms, in order to be able to do an accurate estimate of the value and standard error of ρ;
- The correct specification of the test may depend on moving average component;
- Only tests for a single unit root;

- Does not consider the possibility of the data exhibiting seasonality and, therefore, being necessary to do a seasonal difference;
- Does not consider the possibility of having structural breaks in data.

According to Dickey & Fuller (1979), the value of ρ reveals the nature of the time series y_t where (1) $|\rho| < 1$ corresponds to a stationary series, (2) $\rho = 1$ to a random-walk, and (3) $|\rho| \geq 1$ designates a non-stationary series. So, the null hypotheses relies on this and tests if the series is non-stationary or, by other words, if it has a unit root:

$$\mathrm{H}_0: \rho = 1$$
$$\mathrm{H}_1: \rho < 1$$

To help us interpret all the results, it is important to present the following table provided by Enders (2015):

Table 3.1- Dickey-Fuller Test Summary

Model	**Hypothesis**	**Test Statistic**
$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ
$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ_μ
	$\alpha_0 = \gamma = 0$	ϕ_1
$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \alpha_2 t + \varepsilon_t$	$\gamma = 0$	τ_τ
	$\alpha_0 = \gamma = 0$	ϕ_2
	$\alpha_0 = \gamma = \alpha_2 = 0$	ϕ_3

Source: Enders (2015)

Where the gamma statistics are obtained by:

$$\phi_i = \frac{\left[\frac{SSR(restricted) - SSR(unrestricted)}{r}\right]}{\left[\frac{SSR(unrestricted)}{T-k}\right]} \tag{3.49}$$

Where,

SSR (restricted): The sum of the squared residuals from the restricted model;

SSR (unrestricted): The sum of the squared residuals from the unrestricted model;

R: Number of restrictions;

T: Number of observations;

K: Number of parameters estimated in the unrestricted model.

T-k: degrees of freedom in the unrestricted model.

3.4.1.2 PP Test

An alternative to the ADF test is the Phillips-Perron (PP) test proposed by Phillips (1987). In this new test, the null and alternative hypotheses are the same as ADF but, contrary to them, it is assumed that the errors are dependent and heterogeneously distributed. As we have seen before, in economic series the trend may not be stochastic but linear/deterministic, and it is frequently detected the existence of a drift. So, Phillips& Perron (1988) expanded the concept of the PP test and included the cases where we have a random walk with drift and/or a random walk with drift and a linear trend.

3.4.1.3 KPSS Test

Kwiatkowski, Phillips, Schmidt, & Shin (1992) represent a time series as a set of three components (deterministic trend, a random walk, and a stationary error) and propose the stationarity test KPSS. This test takes a distinct approach and treats the null and alternative hypotheses differently, defining in the opposite way of the previous two tests. Instead of considering that the null hypothesis states that the series is non-stationary, the KPSS considers the absence of unit root. So, the null hypothesis corresponds to stationarity, i.e., the series is not a random walk.

$$\mathrm{H_0}\text{: Variance of the random walk} = 0$$

$$\mathrm{H_1}\text{: Variance of the random walk} \neq 0$$

It is important to perform these two types of tests because, as the hypothesis that they test are different, besides trying to conclude if the series is stationary, we can observe the cases where the lack of information does not allow us to conclude about this characteristic.

3.4.2 Normality Tests

Checking the normality of errors is crucial to some methods, which depend on the confirmation of this assumption to guarantee its validity. This motivated Shapiro & Wilk (1965) to develop the Shapiro-wilk test. The main problem of this test is that it is more recommended for small samples, not being the most accurate when we have a large sample.

An alternative to the Shapiro-wilk test is the Kolmogorov-Smirnov test proposed by Lilliefors (1967), which is more convenient in cases for which we do not know the mean and variance of our process.

Besides these two tests, Jarque & Bera (1987) proposed one of the most popular tests to conclude about the normality of the errors. This test is known for its asymptotic validity and the incorporation of the coefficients of skewness (b_1) and kurtosis (b_2). The aim of the test is defined by the null hypothesis, which states that the errors are normally distributed, and test statistics is given by:

$$JB = n\left[\frac{b_1}{6} + \frac{(b_2 - 3)^2}{24}\right] \tag{3.50}$$

3.4.3 Autocorrelation Diagnostic

According to Hyndman & Athanasopoulos (2018), the residuals are the result of the difference between the fitted values (estimates) and the real observations, and the analysis of these are crucial to understand if the process was adequately estimated, using all information possible to enrich the model. To this to happen, the residuals must show no autocorrelation and zero mean. Despite these characteristics, there are two more that are convenient to check, although it does not call into question the precision of our study: constant variance and normal distribution of the residuals. To check the lack of autocorrelation we can resort to the analysis of the autocorrelation function (ACF), or perform the portmanteau tests. Through the ACF we confirm the residuals are white noise if the autocorrelation values are near zero or if 95% of the ACF spikes are within the levels of significance outlined by blue dashed lines, which correspond to $\pm 2/\sqrt{T}$.

Regarding the portmanteau tests, we will consider the Box-Pierce test proposed by Box & Pierce (1970), with the following test-statistic:

$$BP = n\sum_{k=1}^{m} r_k^2 \sim x_{(m)}^2 \tag{3.51}$$

Where n corresponds to the number of observations, m to the maximum lag, and r_k is the autocorrelation coefficient at lag k defined considering uncorrelated random deviates α:

$$r_k = \frac{\sum \alpha_t \alpha_{t-k}}{\sum \alpha_t^2} \tag{3.52}$$

The null hypothesis states that the autocorrelation coefficients are zero, meaning that the series exbibit a random behavior:

$$\mathrm{H}_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0$$
$$\mathrm{H}_1: \exists \rho_m \neq 0$$

To face some problems generated by the use of Box-Pierce test, Ljung & Box (1978) suggested a new test defined by:

$$LB = n(n+2) \sum_{k=1}^{m} \frac{r_k^2}{(n-k)} \sim x_{(m)}^2 \tag{3.53}$$

3.4.4 Conditional Heteroskedasticity

The result of the a conditional heteroscedasticity test is crucial to confirm that we need the ARCH family. We can do this by applying the Ljung-Box test to the square of the residuals, or by performing the ARCH LM test proposed by Engle (1982). Let the white noise error process is represented by μ_t, the null and alternative hypotheses of the latter are described as:

$$\mathrm{H}_0: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$$
$$\mathrm{H}_1: e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_p e_{t-p}^2 + \mu_t$$

3.4.5 Sign Bias test

We can check if the squared standardized residuals are i.i.d by applying the sign bias, negative sign bias and positive sign bias tests. The sign bias test comprises the effect of all shocks, either positive or negative, while the two remaining tests aim to make a distinction. The negative sign bias test focus only on the negative shocks and, as the name already predicts, the positive sign bias test centers on the impact of the positive shocks on volatility. Besides these tests, we can plot the news impact curve, which demonstrates how the volatility of today is affected by a shock that occurred yesterday (Engle & Ng, 1991).

3.4.6 Information Criteria

The optimal orders of p and q to fit the data have a negative impact on the Sum of Squares of the estimated Residuals (SSR) and degrees of freedom, being difficult to find sometimes. As the number of parameters p and q increase, there is an increased number of coefficients to be estimated. On the other hand, we can observe a decrease in the SSR and a loss of degrees of freedom. In order to understand to what extent these consequences compensate, the two most used measures are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), also referred to as Schwarz Information Criterion (SIC). The main difference between them is that the AIC is more suitable for small samples while the BIC is more appropriate to large samples. This means that the BIC is a more parsimonious criteria, with asymptotical value, giving more truth wordy results, while the AIC may tend to choose models with an excess of parameterization. Let T represents the number of observations, L the maximized value of the likelihood function, and n the number of parameters (q + p + constant term), these two criteria can be described as (Enders, 2015):

$$AIC = Tlog(SSR) + 2n = -2\ln(L) + 2n \tag{3.54}$$

$$BIC = Tlog(SSR) + nlog(SSR) = -2ln(L) + nln(T) \tag{3.55}$$

3.4.7 Loss Functions

The procedure of separating the data into two different groups may generate doubts regarding the effectiveness of each model. Therefore, in order to conclude which is the most accurate model in forecasting out-of-sample, we need to resort to some loss functions.

These measures consider the unpredictable part of the observation and can belong to different classes of measures: the scale-dependent errors, the percentage errors, and scaled errors. Within the scale-dependent errors, we have the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE), which are less flexible measures and do not allow to make comparisons between different series. This disadvantage is approached by percentage errors, which already allows to make comparisons between different series. However, it also entails some disadvantages: the measure is not valid when the forecasted value is zero (being infinite or undefined), it is only accurate on a ratio scale, and they are more affected by negative errors than on positive errors. Regarding this class of models, we will consider the Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE). Since these measures rely on the forecast errors e_t, obtained by the difference between the value from forecast y_t and its real value $\hat{y}_t$, we are going to choose the model with the lower value (Hyndman & Athanasopoulos, 2018).

3.4.7.1 Mean Absolute Error

$$MAE = \frac{\sum_{t=1}^{T}|e_t|}{T} = \frac{\sum_{t=1}^{T}|y_t - \hat{y}_t|}{T} \tag{3.56}$$

3.4.7.2 Mean Square Error

$$MSE = \frac{\sum_{t=1}^{T} e_t^2}{T} = \frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T} \tag{3.57}$$

3.4.7.3 Root Mean Square Error

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T} e_t^2}{T}} = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T}} \tag{3.58}$$

3.4.7.4 Mean Percentage Error

$$MPE = \frac{\sum_{t=1}^{T}\left(\frac{y_t - \hat{y}_t}{y_t}\right) * 100}{T} \tag{3.59}$$

3.4.7.5 Mean Absolute Percentage Error

$$MAPE = \frac{\sum_{t=1}^{T}\frac{|y_t - \hat{y}_t|}{y_t}}{T} = \frac{\sum_{t=1}^{T}\frac{|e_t|}{y_t}}{T} \tag{3.60}$$

Where T correspond to the number of observations.

3.4.8 Granger Causality Test

As the VAR is a bi-directional model, its use only makes sense if the two-time series have some kind of influence on each other. To see if this happens, we will apply the granger causality Granger (1969). By using this test, we can conclude if a stationary series X_t Granger causes a stationary series Y_t, and vice versa:

$$X_t = \sum_{j=1}^{m} \alpha_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t \tag{3.61}$$

$$Y_t = \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \eta_t \tag{3.62}$$

Where ε_t and η_t are white noise series with no correlation, and the α_j, b_j, c_j, and d_j are estimated coefficients.

The coefficients that will dictate whether the series are bidirectional, that is, both series causes each other, are the b_j and c_j. Therefore, if (1) $b_j > 0$ we say the series Y_t Granger causes X_t, and if (2) $c_j > 0$, the series X_t Granger causes Y_t, revealing a feedback relationship.

4 Empirical Study

4.1 Stock Index Prices

4.1.1 Stock Index Prices Description

As we already said previously, we will divide our sample in two parts for in-sample and out-of-sample analysis. Our sample is comprehended by the period 01/07/2010 and 31/07/2020, making a total of 2539 daily observations. Within this period, we will define our periods as follows:

- In-sample: from 01/07/2010 to 02/07/2018;
- Out-of-sample: from 03/07/2018 to 31/07/2020.

In Figure 4.1 we can assess the graphical representation of the respective time series, with the blue line representing the index of small firms and the grey line in the representation of the large firm index. From the lack of predictability of the series, revealed by the frequent increases and decreases of the prices over time, it is clear the presence of a stochastic trend. The behaviour of both series suggests that we have a non-stationary series. We will confirm this later by using the unit root tests and by plotting the ACF and partial ACF (PACF).
The steepest falls occur in both indexes around the same time, with the index of the small firms appearing to be more sensitive, revealing more intense reactions (higher losses).

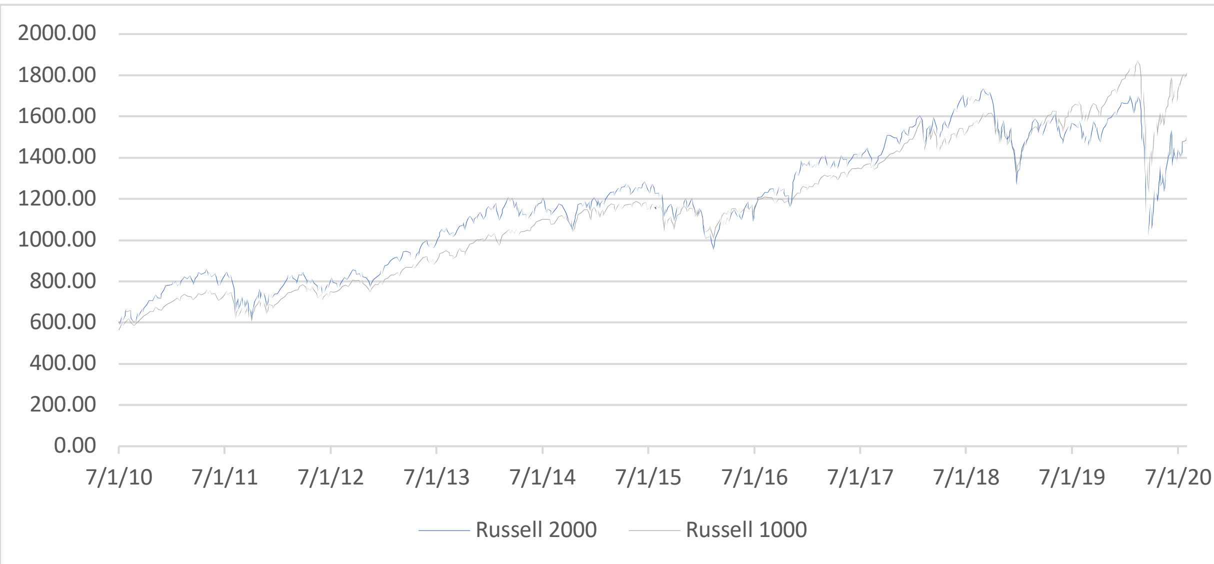

Figure 4.1- Stock price behavior during the last ten years

During the last years, although both indexes present quite similar tendencies, the stock prices of the index that represents the small companies presented higher values than the Russell 1000.

Between 2010 and 2011 both indices showed a growing trend, with the first big drop in 2011. In July 2011 the American stocks started to fall, ending days later the first sharp drop in prices after the 2008 crisis. The American indexes Dow Jones Industrial, S&P500 and Nasdaq Composite registered drops between 5.6% and 6.9%, causing a lot of concern to investors who chose to start selling. The reasons for the loss of more than 6% in early August 2011 was related to the European debt crisis and the fear that the fall in stocks would lead to a possible new American crisis. The prices of the Russell 2000 and Russell 1000 dropped from its maximum of 865.29 and 758.45 on April 29th to 609.49 and 604.42 on September 29th, respectively. After April 29th the stocks of the Russell 2000 recorded a slight increase, registering 858.11 on July 7th , but ended up going down after this period, until October 3rd. The Russell 1000 recorded a similar path, with an increase to 753.32 on July 7th.

After this period, stocks showed a positive evolution until 2015, when a fall of the Chinese currency affected the American stock markets. This event triggered a new crisis, which registered a historic level of volatility in the New York markets, matching the behavior recorded in the "Black Monday" in the 1987 American crisis. This downturn generated a 65% loss in the value of oil and caused stocks to fall worldwide. In the Shanghai composite index, the stocks fell 8.5%, in Europe the drop was quantified in 5%, and in Brazil and Indonesia was recorded a 4% drop. Regarding the United States, the industrial Dow Jones registered a loss of 3.6%, the SP500 of 3.9%, and the Nasdaq composite of 3.8%. On March 21st and March 23rd, it was registered a maximum of 1189.55 on the Russell 1000 and 1295.80 on the Russell 2000. These values fell to 1041.77 and 1083.91 on August 25th and September 29th, respectively.

The behavior observed in 2015 was extended to the next year and at the beginning of 2016 things were not looking very optimistic. The American stock prices fell 10%, leading investors to value other assets such as gold, which ended up appreciating. The main boosters of the fall in stock prices were the historical drop in oil prices, which had not recorded so many losses since 2004, the China slowdown, and the change in interest rate policy. The maximum of the Russell 2000 (1110.44) and Russell 1000 (1116.84) were recorded on January 5th, and fell on February 11th to 953.72 and 1005.89, respectively.

In the upcoming years, the stocks maintained a constant upward trend until 2018 when the trade war between Trump and China, the deceleration of the global economic growth, the increase in interest rates, the Brexit, and the new regulation in the technology sector, stimulated a new crisis. This outbreak had consequences across the globe with the three major United States indices (S&P500, Dow Jones Industrial Average, Nasdaq Composite) dropping from 4% to 6%. Outside the United States, the German index DAX decreased 18%, the French index CAC40 dropped 11%, and the British index FTSE went down 12.5%. The month of December turned out to be the worst Christmas month since the Great Depression. The price of the Russell 2000 and Russell 1000 was 1740.75 on August 31st and 1624.28 on September 20th, ending both to fall on December 24th to 1266.92 and 1298.02, respectively. After this event, the increasing behavior is predominating, with the stock prices of the Russell 1000 index having a sharper growth than the stock prices of the Russell 2000 index. In March, the large firm index started to show prices notoriously higher than small firms and both indices kept a similar growing pattern until 2020 when, almost at the end of our sample, in March 2020, the prices suffer the biggest fall ever, triggered by the COVID-19 pandemic.

This crash generated a high level of concern by the investors that feared we could reach a 20% loss, entering in a bear market for the first time since the 2007-2009 crisis. The bear market was officialized when it was announced the disease COVID-19 was a pandemic, causing serious reactions by the three major American indexes. These indexes recorded a loss of 23% on March 12th, compared with their high in February 2020. In regards to our data, the Russell 2000 had its 2020 maximum on January 16th with 1705.22 and the Russell 1000 on February 19th with 1875.24. After the announcement that the coronavirus fulfilled the requirements to be called a pandemic, the Russell 2000 fell dramatically to 991.16 on March 18th and the Russell 1000 to 1224.45 on March 23rd. From a percentage point of view, there was a loss of 41.97% and 34.70% in the Russell 2000 and Russell 1000, respectively. After this severe period, the prices show high oscillations keeping, however, a growing trend.

4.1.2 Unit Root Tests

We can check the stationarity of the series either by analyzing the ACF of the series, or by performing the unit root tests. Based on the ACF, we conclude the series is non-stationary if exists any dependency of the series (revealed by the slow decay to zero of the autocorrelations spikes) (Hyndman & Athanasopoulos, 2018).

To compute this type of tests we considered the "urca" package in the program R. To perform the ADF test we used the "ur.df" function, which was specified as follows:

- "type": the test regression can assume three different forms: (1) "none" (there is no intercept and no trend), (2) "drift" (there is an intercept), and (3) "trend" (there are an intercept and a trend). We considered these three forms;
- "lag": in order help us define the optimal lag length to our model we considered the BIC information criteria by using the "selectlags" part.

Table 4.1- Augmented Dickey-Fuller Test applied to Russell 2000 stock prices

Type	**Test Statistic**	**Lags**	**Significance levels**			**Value of the test-statistic**
			10pct	**5pct**	**1pct**	
None	tau1	7	-1.62	-1.95	-2.58	0.5503
Drift	tau2	7	-2.57	-2.86	-3.43	-1.7397
	phi1		3.78	4.59	6.43	2.0235
Trend	tau3	7	-3.12	-3.41	-3.96	- 3.4543
	phi2		4.03	4.68	6.09	4.394
	phi3		5.34	6.25	8.27	6.0791

Table 4.2- Augmented Dickey-Fuller Test applied to Russell 1000 stock prices

Type	**Test Statistic**	**Lags**	**Significance levels**			**Value of the test-statistic**
			10pct	**5pct**	**1pct**	
None	tau1	9	-1.62	-1.95	-2.58	1.5265
Drift	tau2	9	-2.57	-2.86	-3.43	-0.6673
	phi1		3.78	4.59	6.43	1.8164
Trend	tau3	9	-3.12	-3.41	-3.96	-4.9402
	phi2		4.03	4.68	6.09	9.2422
	phi3		5.34	6.25	8.27	12.2548

Regarding the test regression with an intercept and a trend, we can conclude about the presence of a unit root in the series by comparing the value of test-statistics to the τ_{τ} (tau3 in the output) significance levels. Starting by analysing the Russell 2000, the results are a little bit odd, as the value of the test-statistic leads us to different conclusions. The value of the test-statistic (-3.4543) is lower at the significance levels 10% (-3.12) and 5% (-3.41), leading us to reject the null hypotheses and conclude that the series presents a stationary behavior. At the significance level 1% (-3.96) the value of the test statistic is higher, leading to the non-rejection the null hypotheses, meaning that the series has a unit root. The problem with this version of the test is that the series may not have a drift and/or a trend and, therefore, the tau3 may give unreliable results. So, we rely on the ϕ_3 (phi3 in the output) to test if α_0, γ and α_2 equal zero (see Table 3.1). As the value of the test-statistic (6.0791) is lower than the critical significance levels at 5% (6.25) and 1% (8.27), we do not reject the null and conclude that the series has a unit root but does not have a deterministic trend. By contrary, at the significance level 10% (5.34) we conclude that, as the value of the test-statistic is higher, at least one of the coefficients is different from zero.

As it does not include a trend to some significance levels, we proceed to the analysis of type "drift" (test regression with intercept). After analyzing the value of the test-statistic (-1.7397), we can observe that it is higher than the τ_μ (tau2 in the output) significance levels. So, we do not reject the null and conclude that the series is non-stationary. As we did to the type "trend", we rely on the ϕ_1 (phi1 in the output) and conclude that, as the value of the test-statistic (2.0235) is lower than the critical values, we do not reject the null and conclude that the series has a unit root but does not have a drift.

As it was confirmed that there is not a drift nor a trend, the most adequate way to check the stationarity of the series is by relying on τ (tau1 in the output). As the value of the test-statistic (0.5503) is higher than the critical values, we do not reject the null hypothesis and, therefore, the evidence points to the existence of a unit root.

In terms of the index of large firms, the value of the test statistic of the τ_τ (-4.9402) is lower than all significance levels and, therefore, we reject the null. Following the same line of thought the index of small firms, we need to look at the ϕ_3 to see if this type of regression is the most trustable. As the value of the test-statistic (12.2548) is higher than all critical values (for different significance levels), we should reject the null hypothesis, meaning that at least one of the coefficients is different from zero. The same happens to the ϕ_2 (phi2 in the output) significance levels. Given these results, we cannot be sure if the series has a deterministic trend nor a drift, and if the results given by τ_τ are reliable. To understand this, we will evaluate the type "drift". Based on the value of the test statistic (-0.6673) we do not reject the null and conclude that the series has a unit root. However, after looking at the significance levels of ϕ_1, as these are higher than the value of the test-statistic (1.8164), we do not reject the null and conclude that the series has a unit root and no drift. So, the most adequate way to check the stationarity of the series is by analyzing the τ. In this case, as the value of the test-statistic (1.5265) is higher than the critical values, we do not reject the null hypothesis and can affirm that the series has a unit root.

In conclusion, we confirm that both Russell 2000 and Russell 1000 series is a random walk and that it has non-stationary behavior.

To run the PP test, we considered the "ur.pp" function, settled as:

- "type": we considered the "Z.tau" type;
- "model": we can chose between "constant" and "trend", in order to include the deterministic component of the test regression (that is, test the cases whether there is a trend or not). We only considered the constant model, as we saw in the ADF test that the trend model may not be the best;
- "lags": reveals the number of lags considered for the error term correction. We performed both "short" and "long" lags.

Table 4.3- Phillips-Perron Test applied to Russell 2000 stock prices

Type	**Model**	**Lags**	**Significance levels**			**Value of the test-statistic**
			10pct	**5pct**	**1pct**	
Z-tau	Constant	Short	-2.567667	-2.86318	-3.435868	-1.7807
		Long				-1.8036

Table 4.4- Phillips-Perron Test applied to Russell 1000 stock prices

Type	**Model**	**Lags**	**Significance levels**			**Value of the test-statistic**
			10pct	**5pct**	**1pct**	
Z-tau	Constant	Short	-2.567668	-2.863181	-3.435869	-0.6498
		Long				-0.6735

The value of the test-statistics of the Russell 2000 (-1.7807 and -1.8036) and Russell 1000 (-0.6498 and -0.6735) are higher than the critical values to both short and long lags. This implies the non-rejection of the null hypothesis and confirms the presence of a unit root in both indexes.

Regarding the KPSS, the function used in R to run this stationary test was the "ur.kpss":

- "type": is related to the deterministic part of the series, being considered "mu" when the deterministic part is constant, and "tau" when is constant with a linear trend;
- "lags": corresponds to the maximum number of lags used for the correction of the error term. This can be obtained by $\sqrt[4]{4 * (n/100)}$ or $\sqrt[4]{12 * (n/100)}$, depending on if we choose "short" or "long". On contrary, if the option "nil" is used, the error term is not corrected.

Table 4.5- Kwiatkowski-Phillips-Schmidt-Shin Test applied to Russell 2000 stock prices

Type	**Lags**	**Significance levels**				**Value of the test-statistic**
		10pct	**5pct**	**2.5pct**	**1pct**	
Tau	Short	0.119	0.146	0.176	0.216	1.1328
	Long					0.4031
	Nil					9.9047
Mu	Short	0.347	0.463	0.574	0.739	25.1489
	Long					8.4977
	Nil					224.8546

Table 4.6- Kwiatkowski-Phillips-Schmidt-Shin Test applied to Russell 1000 stock prices

Type	**Lags**	**Significance levels**				**Value of the test-statistic**
		10pct	**5pct**	**2.5pct**	**1pct**	
Tau	Short	0.119	0.146	0.176	0.216	0.6853
	Long					0.2563
	Nil					5.8589
Mu	Short	0.347	0.463	0.574	0.739	27.2502
	Long					9.1938
	Nil					243.7732

As we can observe, the value of the test-statistic is higher than the critical values obtained for the four different significance levels (and to the different types and lags). Therefore, we reject the null hypotheses and the series is non-stationary.

4.1.3 Autocorrelation and Partial Autocorrelation Functions

As we said previously, the behaviour of a random walk series is also reflected in its ACF and PACF. If we plot these two functions for both Russell 2000 and Russell 1000 we can confirm this lack of stationarity. According to Figure 4.2 and Figure 4.3, both indexes show an ACF with slow decay to zero and, therefore, a high degree of persistence in autocorrelation lags is confirmed. This leads us to confirm the previous results and conclude that the data has non-stationary behaviour.

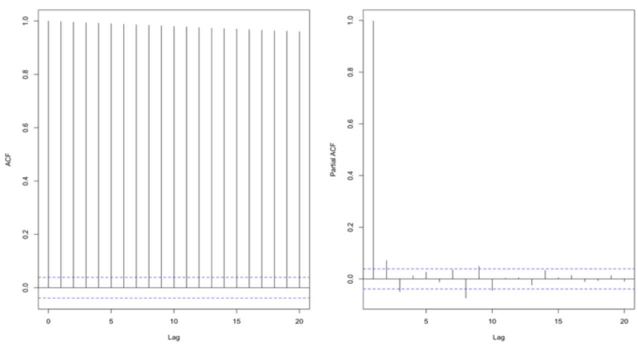

Figure 4.2- Autocorrelation and Partial Autocorrelation Functions of the Russell 2000 Stock Prices

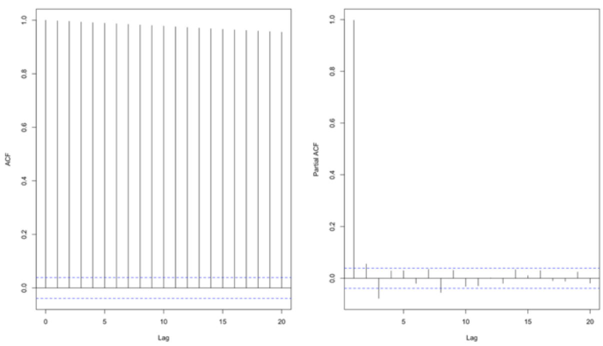

Figure 4.3- Autocorrelation and Partial Autocorrelation Functions of the Russell 1000 Stock Prices

So, the results obtained in these two sub-sections, reinforce that we have an integrated series and therefore our data is better described by the ARIMA conditional mean model.

4.2 Stock Index Returns

As we said previously, we need to face the non-stationarity identified through the unit root tests. To do this, we consider the continuously compounded returns (r_t) for the indexes Russell 2000 and Russell 1000:

$$r_t = \ln(P_t) - \ln(P_{t-1}) = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{4.1}$$

Where P_t are the prices of the indexes at time t ($t = 0, \dots, t = 2539$).

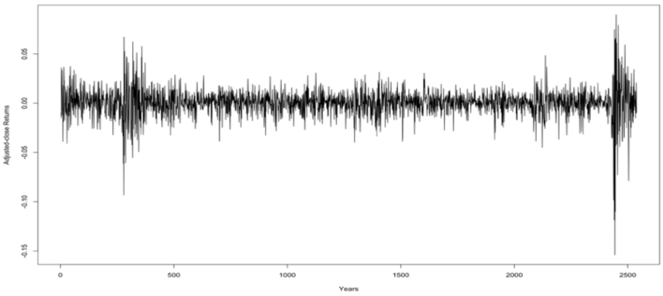

Figure 4.4- Russell 2000 Continuously Compounded Returns

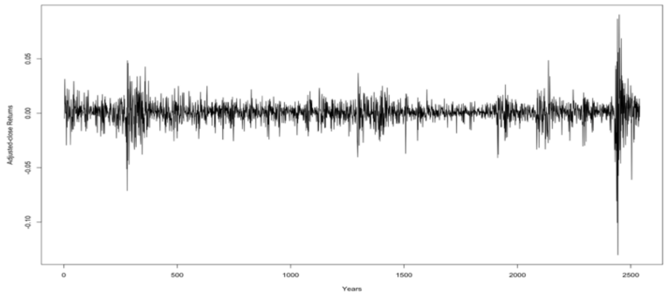

Figure 4.5- Russell 1000 Continuously Compounded Returns

As we can see in Figure 4.4 and Figure 4.5, after considering the continuously compounded returns, the series does not show a stochastic trend anymore and it is clearly observable the existence of the stylized fact volatility clustering. If we look more carefully to the graphs we can see that the index that represents the small capitalization firms oscillates more than the one from large firms. This means that in periods of higher volatility, the small firms show an increased response, relatively to higher firms. By the plot of squared and absolute returns we can confirm the persistence, revealed by the existence of clusters:

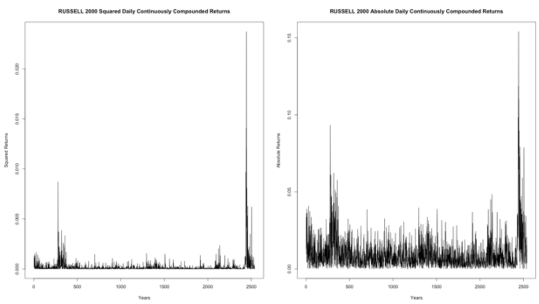

Figure 4.6- Squared and Absolute Daily Returns of Russell 2000

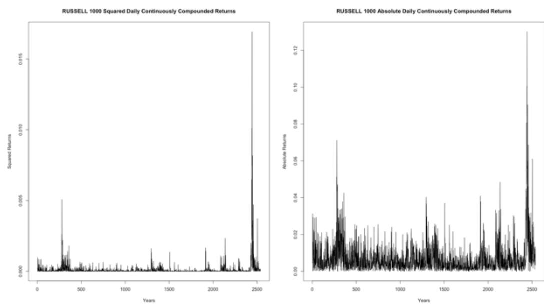

Figure 4.7- Squared and Absolute Daily Returns of Russell 1000

Table 4.7- Stock returns descriptive statistics

	Minimum	**25th Quartile**	**Median**	**Mean**	**75th Quartile**	**Maximum**
Russell 2000	-0.15399	-0.00562	0.00094	0.00035	0.00747	0.08976
Russell 1000	-0.13010	-0.00341	0.00066	0.00046	0.00524	0.09041

Table 4.7 contains the descriptive statistic of the continuously compounded returns, where we can confirm that the mean is now around zero.

4.2.1 Normality

By the analysis of the QQ plots, we see that the dots clearly distanced themselves from the line, indicating that none of the indexes is normally distributed. In regards to the histograms, we see there is a slight deviation from normality but the graphs are not very clear.

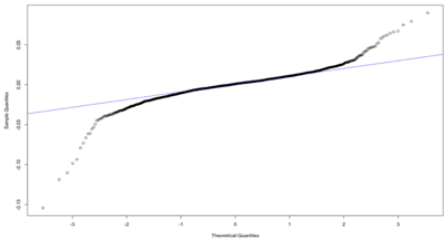

Figure 4.8- Russell 2000 Stock Returns QQ plot

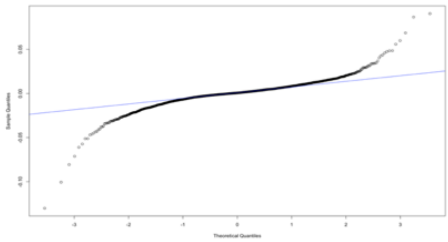

Figure 4.9- Russell 1000 Stock Returns QQ plot

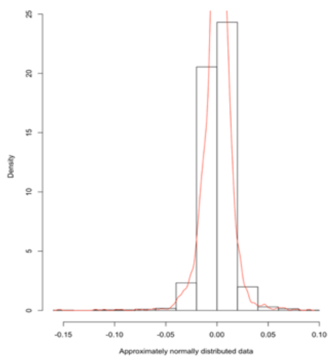

Figure 4.10- Russell 2000 Stock Returns Histogram with the normal distribution

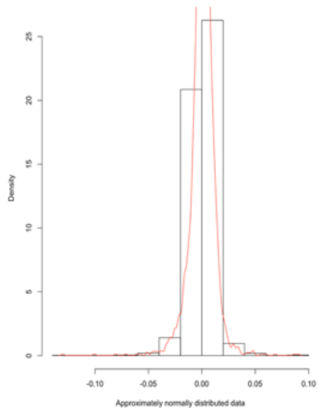

Figure 4.11- Russell 1000 Stock Returns Histogram with the normal distribution

To clarify, we present some tests to confirm the lack of normality in Table 4.8:

Table 4.8- Normality Tests applied to Stock Returns

	Russell 2000	**Russell 1000**
Jarque-Bera Test		
χ^2	19763	35933
p-value	< 2.2e-16	< 2.2e-16
Kolmogorov-Smirnov Test		
w	0.088581	0.11805
p-value	< 2.2e-16	< 2.2e-16
Skewness Coefficient	-1.037408	-0.9855515
Kurtosis Coefficient	16.51229	21.32764

By relying on the Jarque-Bera and Kolmogorov-Smirnov tests, we conclude that the data is not normally distributed, as the p-value is lower than the significance level 0.05 to all three tests. We do not present results for the Shapiro-Wilk's test because it is not very relevant in our study as it most appropriate to small samples. Regarding the skewness and kurtosis coefficients, the data is considered normal if these coefficients equal zero and three, respectively. In our data, the coefficient of skewness is negative for both indexes, revealing that the data is negatively skewed, or skewed to the left. In terms of the coefficient of kurtosis, as the value of the coefficient is higher than 3 to both small and large firms, the data is leptokurtic or, by other words, fat tailed, meaning that it has more observations in the tails. Contrary to Chelley-Steeley & Steeley (1996), the small firms present less leptokurtosis than large firms.

4.2.2 Unit Root Tests

As we can see in ADF test (Annex A and Annex B), PP test (Annex C and Annex D), and KPSS test (Annex E and Annex F) the evidence points to the absence of a unit root and therefore differencing the series solved the non-stationary problem.

4.3 Conditional Mean Model

4.3.1 Box-Jenkins Methodology

To help us determine the estimates for the parameters of the conditional mean model, we will consider the Box-Jenkins methodology. This method is constituted by three different phases: identification, estimation, and the diagnostic of the residuals. In the first phase, the aim is to identify the most adequate process to fit our data (that is, find the values for the orders for the ARIMA process).According to Anderson (1977), the order of q and p are obtained by analyzing the ACF and PACF, respectively, for the first k lags (where k corresponds to the minimum between 20 and n/4). The second and third steps consist of estimating the coefficients of the process and doing a diagnostic of the residuals to confirm that they are white noise, respectively (Enders, 2015). Both of these functions aim to evaluate the relationship between two observations with the difference that, in the ACF, we consider the effect of the lags (Hyndman & Athanasopoulos, 2018).

As we can see in Figure 4.12 and Figure 4.13, the estimates for the coefficients do not converge to zero at any lag, being difficult to conclude about the optimal order of the process through the analysis of these graphs:

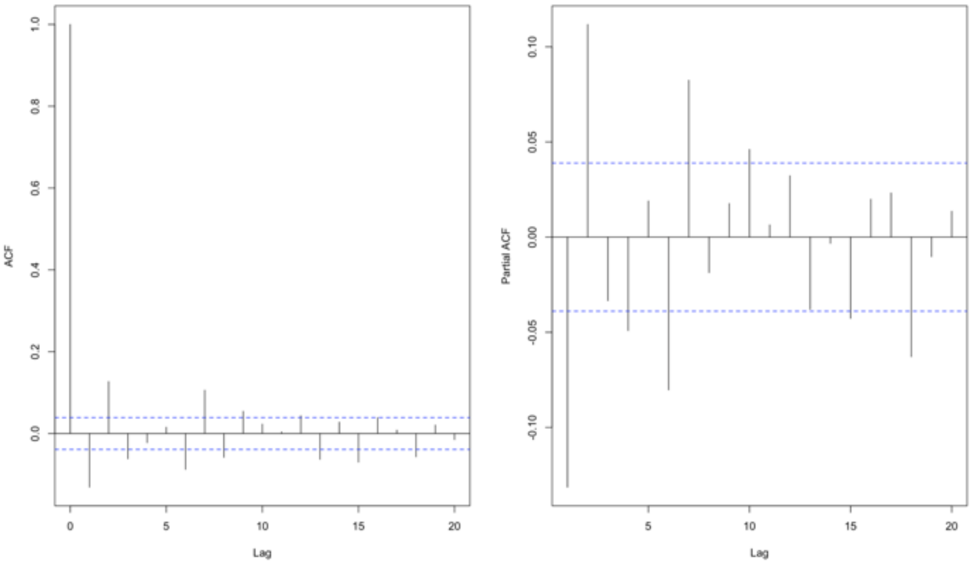

Figure 4.12- Autocorrelation and Partial Autocorrelation Functions of the Russell 2000 Stock Returns

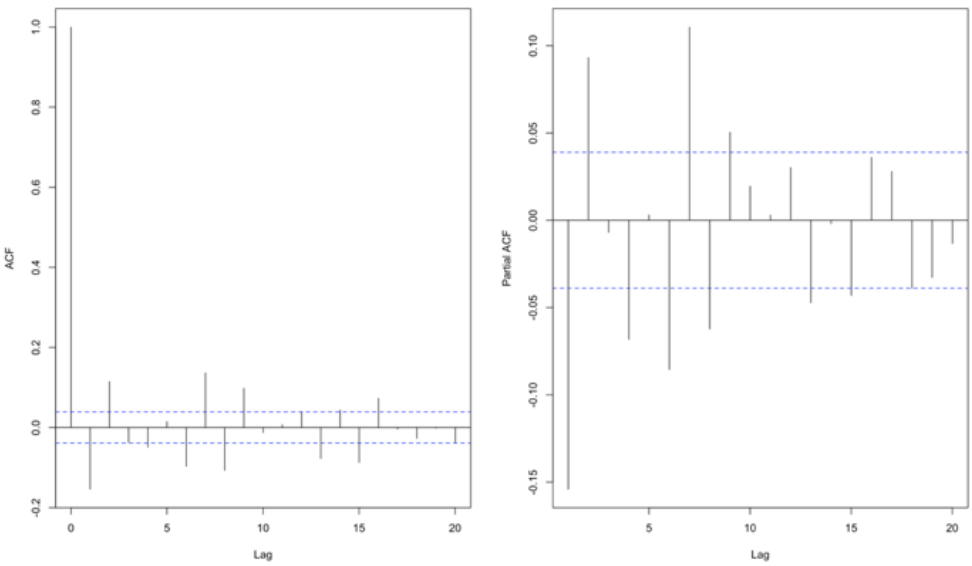

Figure 4.13- Autocorrelation and Partial Autocorrelation Functions of the Russell 1000 Stock Returns

So, in order to find the best orders, we followed the same line of thought of Maganlal (2019) and estimated the model until the orders p=3 and q=3, considering d=0, as we know *a piori* that our process is better characterized by an ARMA, since we already computed the continuously compounded returns.

Table 4.9- Information Criteria of ARMA (p,0,q) Models

	Russell 2000		Russell 1000	
Orders (p, d, q)	**AIC**	**BIC**	**AIC**	**BIC**
(0,0,0)	-14 379.88	-14 368.2	-15 649.88	-15 638.2
(0,0,1)	-14 413.9	-14 396.38	-15 698.58	-15 681.06
(0,0,2)	-14 447.25	-14 423.89	-15 727.9	-15 704.54
(0,0,3)	-14 456.48	-14 427.29	-15 731.54	-15 702.35
(1,0,0)	-14 422.12	-14 404.6	-15 708.77	-15 691.26
(1,0,1)	-14 442.91	-14 419.55	-15 721.96	-15 698.61
(1,0,2)	-14 453.84	-14 424.64	-15 729.25	-15 700.06
(1,0,3)	-14 454.49	-14 419.46	-15 734.07	-15 699.04
(2,0,0)	-14 452	-14 428.64	-15 728.93	-15 705.58
(2,0,1)	-14 451.43	-14 422.23	-15 726.97	-15 697.78
(2,0,2)	-14 492.18	-14 457.15	-15 792.68	-15 757.65
(2,0,3)	-14 491.16	-14 450.29	-15 733.93	-15 693.06
(3,0,0)	-14 452.85	-14 423.65	-15 727.05	-15 697.86
(3,0,1)	-14 454.48	-14 419.44	-15 729.6	-15 694.56
(3,0,2)	-14 454.11	-14 413.23	-15 736.1	-15 695.23
(3,0,3)	-14 453.47	-14 406.76	-15 734.66	-15 687.95

Note: the highlighted results correspond to the lowest information criteria values.

Based on both information criteria, the best process to define both indexes is an ARMA with autoregressive order equal to 2 and moving average order equal to 2, that is, an ARIMA (2,0,2).

4.3.2 Autocorrelation tests

As we said formerly, we need to do a diagnostic of the residuals and check if they are generated by a white noise process and, therefore, conclude if it is in accordance with the Efficient Market Hypothesis. To check if the residuals are autocorrelated we consider the Box-Pierce and Ljung-Box portmanteau tests, obtained by performing the function “box.test”. According with the documentation of the test, the number of lags should be higher than “fitdf”, which corresponds to the sum of the autoregressive and moving average orders of the ARIMA process. Hyndman & Athanasopoulos (2018) suggests that we should consider 10 lags or, if that is too large, one fifth of the number of observations in our series. So, based on this, we considered ten lag.

Table 4.10- Box-Pierce and Ljung-Box Tests applied to the residuals

	Russell 2000	Russell 1000
Box-Pierce Test		
χ^2	14.873	21.254
p-value	0.02127	0.001651
Ljung-Box Test		
χ^2	14.924	21.32
p-value	0.02086	0.001607

As the p-value is lower than the significance level 5%, we should reject the null hypothesis and conclude that there is some autocorrelation and, therefore, the process was not generated by a white noise process. Contrary to some studies already presented in the literature review, the index constituted by large firms seems to present more autocorrelation than the index of small firms.

If we plot the ACF of residuals and square residuals, we observe some interesting properties. By the ACF of the residuals, represented by Figure 4.14 and Figure 4.15, we can confirm that the residuals are not autocorrelated, as some of the autocorrelation values are outside the significance level. This conclusion is contrary to what we were expecting but, in some cases, we cannot eliminate 100% of the autocorrelation.

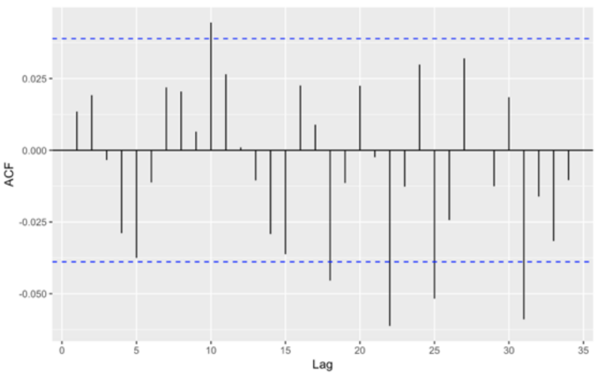

Figure 4.14- ACF Applied to the Residuals of the Russell 2000 ARMA Model

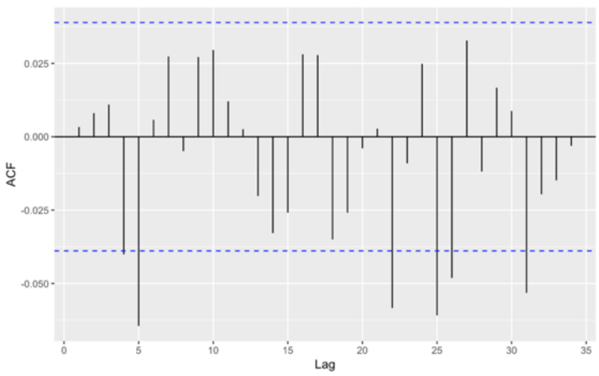

Figure 4.15- ACF Applied to the Residuals of the Russell 1000 ARMA Model

By the ACF of the squared residuals (Figure 4.16 and Figure 4.17), we see that, by its slow decay to zero, the residuals are not linearly independent and therefore, the ergodicity property is not confirmed. This suggests the existence of conditional heteroscedasticity:

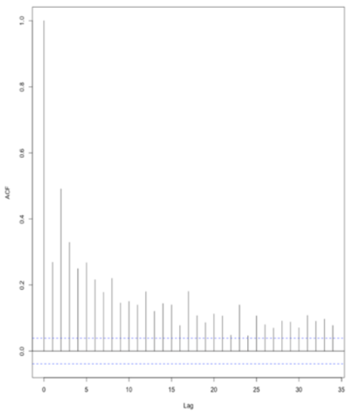

Figure 4.16- ACF of the Russell 2000 Squared Residuals

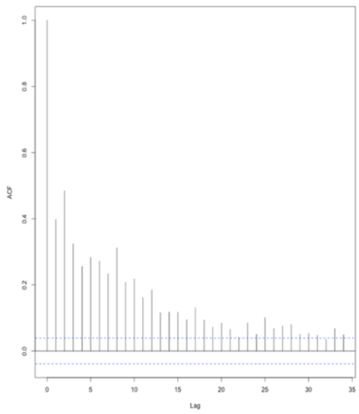

Figure 4.17- ACF of the Russell 1000 Squared Residuals

Regarding normality, we can see by the histograms (Figure 4.18 and Figure 4.19) and by the tests on Annex G, that the residuals are slightly skewed and, therefore, do not follow a normal distribution:

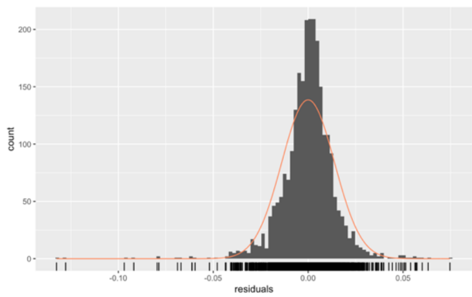

Figure 4.18- Russell 2000 Residuals Histogram

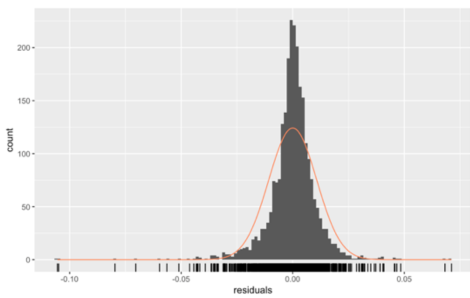

Figure 4.19- Russell 1000 Residuals Histogram

4.3.3 Conditional Heteroskedasticity

What will dictate whether it is necessary or not to use the conditional heteroskedastic models is the existence of conditional heteroscedasticity. Although we are practically certain that the data display this ARCH effect, as the conditional mean model does not deal with this problem, it is crucial to perform some tests to check it. To do this, we used the function “arch.test” from the “aTSA” package, and applied it to the residuals of the ARIMA(2,0,2) model:

Table 4.11- Lagrange Multiplier Test

Order	Russell 2000		Russell 1000	
	LM	p-value	LM	p-value
4	1094	0	1143	0
8	534	0	523	0
12	349	0	329	0
16	251	0	239	0
20	190	0	189	0
24	153	0	155	0

Based on the Lagrange Multiplier test, as the p-value is lower than the significance level 5%, we should reject the null hypothesis and affirm that there exists conditional heteroscedasticity or ARCH effect.

Table 4.12- Ljung-Box Tests applied to the squared residuals

	Russell 2000	Russell 1000
Ljung-Box Test		
χ^2	1844	2477.4
p-value	< 2.2e-16	< 2.2e-16

According to the Ljung-Box test applied the square of residuals, as the p-value is lower than the significance level 5%, we should reject the null hypotheses, confirming the result obtained through the Lagrange Multiplier test. Therefore, we should proceed with the estimation of the conditional variance models.

4.4 Univariate Conditional Variance Models

4.4.1 In-sample analysis

To estimate the conditional variance models we used the “ugarchspec”, “ugarchfit” and “ugarchforecast” functions of the rugarch package. The "ugarchspec" allows to specify the model we want to estimate by choosing the ARMA model, conditional variance model, and type of distribution. The model distribution is specified in the "distribution.model" part and, within the different distributions we can choose, we opted to estimate the models for the normal distribution, student-t distribution, and GED distribution. Concerning the models themselves, they are defined in the "model" part and, in line with what we defined in the methodology section, we will consider: “sGARCH”, “eGARCH”, “gjrGARCH”, “apARCH” and, within the group “fGARCH”, the “TGARCH”. The “ugarchfit” is used to fit the model and we only defined the "spec" (where we selected the model specified before), and the "data" where we considered the in-sample period. In the “solver” option, we chose the “hybrid” since it is the safest choice. By using this type of solver, we avoid having to define additional characteristics, such as the "solver.control" and "fit.control", decreasing the probability of error in the estimation of the models. In short, the hybrid solver starts by estimating the model by the “solnp” solver and re-estimates it (using the “nlminb”, “gosolnp”, and “nloptr” options) until the one that best matches the model is found, ignoring the cases where convergence problems occur (Lin, 2019). Lastly, we used the “ugarchforecast” to make the out-of-sample forecasts.

Regarding the orders of the models, we took into account the Bollerslev, Chou, & Kroner (1992) article, and chose $p = q = 1$, as they claim that it might be sufficient to obtain a good estimation:

Table 4.13- GARCH (1,1) model estimates (the results in brackets represent the p-values)

	Normal Distribution		Student-t Distribution		GED Distribution	
	Russell 2000	**Russell 1000**	**Russell 2000**	**Russell 1000**	**Russell 2000**	**Russell 1000**
Conditional Mean Estimates						
mu	0.000647 (0.002897)	0.000697 (0.00000)	0.000819 (0.00000)	0.000697 (0.00000)	0.000852 (0.000000)	0.000750 (0.000000)
ar1	-0.468338 (0.000000)	0.491746 (0.00000)	-0.012534 (0.80126)	0.491746 (0.00000)	0.000085 (0.998619)	0.531469 (0.000000)
ar2	-0.912683 (0.000000)	0.436425 (0.00000)	0.866683 (0.00000)	0.436425 (0.00000)	0.862393 (0.000000)	0.372548 (0.000000)
ma1	0.446344 (0.000000)	-0.543935 (0.00000)	-0.029588 (0.49942)	-0.543935 (0.00000)	-0.042562 (0.273051)	-0.599911 (0.000000)
ma2	- 0.902555 (0.000000)	-0.419057 (0.00000)	-0.880137 (0.00000)	-0.419057 (0.00000)	- 0.873460 (0.000000)	-0.341742 (0.000000)
Conditional Variance Estimates						
omega	0.000006 (0.000000)	0.000004 (0.009728)	0.000004 (0.15709)	0.000004 (0.009728)	0.000005 (0.559399)	0.000003 (0.10485)
alpha1	0.105757 (0.000000)	0.158538 (0.000000)	0.083293 (0.000001)	0.158538 (0.00000)	0.095479 (0.000000)	0.157612 (0.000000)
beta1	0.849722 (0.000000)	0.797203 (0.00000)	0.890518 (0.00000)	0.797203 (0.00000)	0.868746 (0.000000)	0.812429 (0.000000)
shape	-	-	8.384303 (0.00000)	-	1.486666 (0.000000)	1.207678 (0.000000)

Note: the highlighted results correspond to the coefficients that are not statistically significant.

The outputs corresponding to the conditional mean model are given by "mu", "ar1", “ar2”, “ma1” and "ma2", which report the intercept, autoregressive, and moving average components. Regarding the conditional variance models, the "omega" corresponds to the constant/intercept, while the "alpha1" and "beta" are ARCH and GARCH terms, respectively. In addition to these, we have the “shape”, which reports the number of degrees of freedom.

As we can see in Table 4.13, almost all the significance tests present p-value lower than 5% being, therefore, all the estimates for the coefficients statistically significant except the first-order autoregressive, first-order moving average for the Russell 2000 under student-t and GED distributions, and the omega coefficient to the Russell 2000 under student-t distribution, and both indexes with GED distributions. This is not problematic since the main focus of our study is on the conditional variance part. Regarding the omega, we do not see this as an obstacle to the study because, if the omega is equal to zero and meet the requirements to ε_t be covariance stationary as we have seen it should be (that is, the sum of alpha1 and beta1 is less than one), the conditional variance of the model will tend to decrease, which is not what we aspire (Lin, 2019).

As we saw in methodology, there are some properties of the GARCH model that we have to confirm if they are observed. The first one is related to the nonnegativity of the coefficients and, as we can see, as the estimates for the parameter alpha1 and beta1 always show positive values, so this condition is satisfied and the model is valid. The second has to do with the sum of alpha and beta and, as this result in a value below 1, the ε_t is covariance stationary.

The estimates of alpha1 report the short-run persistence of shows on conditional volatility, while the beta1 reports the long-run persistence. Based on these, we can see that in the index constituted by small firms the value of alpha1 is lower and the value of beta1 is higher. This means that a shock has a greater immediate impact on large companies but that, in a longer time horizon, it affects more the small companies.

Table 4.14- EGARCH (1,1) model estimates (the results in brackets represent the p-values)

	Normal Distribution		Student-t Distribution		GED Distribution	
	Russell 2000	Russell 1000	Russell 2000	Russell 1000	Russell 2000	Russell 1000
Conditional Mean Estimates						
mu	0.000288 (0.140513)	0.000427 (0.002679)	0.000506 (0.010477)	0.000586 (0.000000)	0.000561 (0.005461)	0.000561 (0.000000)
ar1	-0.980976 (0.000000)	-0.148191 (0.170760)	0.214451 (0.014113)	0.358546 (0.000000)	-1.091724 (0.000000)	0.177616 (0.002461)
ar2	-0.052596 (0.000398)	0.516892 (0.000000)	-0.599847 (0.000000)	0.285929 (0.000000)	-0.151901 (0.000000)	0.570781 (0.000000)
ma1	0.942345 (0.000000)	0.095361 (0.356905)	0.172277 (0.044765)	-0.411449 (0.000000)	1.039518 (0.000000)	-0.229735 (0.000172)
ma2	0.025895 (0.000000)	-0.528563 (0.000000)	0.614169 (0.000000)	-0.262739 (0.000000)	0.111314 (0.000012)	-0.546872 (0.000000)
Conditional Variance Estimates						
omega	-0.278672 (0.000000)	-0.548892 (0.000000)	-0.223436 (0.000000)	-0.466002 (0.000000)	-0.245985 (0.000000)	-0.522222 (0.000049)
alpha1	-0.126142 (0.000000)	-0.179854 (0.000000)	-0.132613 (0.000000)	-0.212433 (0.000000)	-0.125579 (0.000000)	-0.200088 (0.000000)
beta1	0.969012 (0.000000)	0.942808 (0.000000)	0.975735 (0.000000)	0.952734 (0.000000)	0.973307 (0.000000)	0.947054 (0.000000)
gamma1	0.113938 (0.000000)	0.197039 (0.000000)	0.112804 (0.000000)	0.193906 (0.000000)	0.112462 (0.000000)	0.200485 (0.000002)
shape	-	-	9.987331 (0.000001)	5.505072 (0.000000)	1.560709 (0.000000)	1.263061 (0.000000)

Note: the highlighted results correspond to the coefficients that are not statistically significant.

In the EGARCH model, we do not anticipate any problems in terms of the significance of the coefficients. The conditional variance estimates are all statistically significant unlike some conditional mean coefficients, where this changes slightly. This only happens with the normal distribution where the intercept estimate is insignificant for the Russell 2000 and, for the Russell 1000, the insignificant coefficients are the first-order autoregressive and first-order moving average. This is not problematic since removing it would not affect the overall accuracy of the results. Under the student-t and GED distributions, all the estimates are statistically significant.

As we can see in Table 4.14, the estimate for the parameter alpha1 assumes negative values, which confirms the existence of the leverage effect, and based on this coefficient, it seems that larger firms are more affected by the arrival of bad news. The additional coefficient gamma1 is a specific coefficient of this model, which represents the asymmetry component of the series and confirms the existence of the leverage effect by its nonnegativity.

The results of beta1 meet the same conclusion of the GARCH model: shocks to small firms are more persistent than to large firms.

The alpha1 results are quite surprising so we verified if they have any validity by checking if the difference of the estimates is statistically significant with the following procedure:

$$-2 < \frac{\hat{\alpha}_{Russell\ 2000} - \hat{\alpha}_{Russell\ 1000}}{\sqrt{(\sigma_{Russell\ 2000})^2 + (\sigma_{Russell\ 1000})^2}} > 2 \tag{4.2}$$

As the values are all above 2, the difference is statistically significant and the results given by alpha1 are reliable.

Table 4.15- GJR-GARCH (1,1) model estimates (the results in brackets represent the p-values)

	Normal Distribution		**Student-t Distribution**		**GED Distribution**	
	Russell 2000	**Russell 1000**	**Russell 2000**	**Russell 1000**	**Russell 2000**	**Russell 1000**
Conditional Mean Estimates						
mu	0.000391 (0.058465)	0.000476 (0.000004)	0.000622 (0.001376)	0.000675 (0.000000)	0.000572 (0.006401)	0.000623 (0.000000)
ar1	-0.248721 (0.357339)	0.212278 (0.000000)	-0.220820 (0.334866)	0.386037 (0.000000)	-0.109418 (0.435313)	0.246980 (0.000000)
ar2	0.635211 (0.003855)	0.742587 (0.000000)	0.668730 (0.000744)	0.525178 (0.000000)	-0.596839 (0.000000)	0.658794 (0.000000)
ma1	0.212171 (0.423028)	-0.247063 (0.000000)	0.180176 (0.419134)	-0.435422 (0.000000)	0.068594 (0.614770)	-0.298505 (0.000000)
ma2	-0.649303 (0.001815)	-0.724282 (0.000000)	-0.684677 (0.000253)	-0.503287 (0.000000)	0.620751 (0.000000)	-0.634562 (0.000000)
Conditional Variance Estimates						
omega	0.000005 (0.00000)	0.000004 (0.000000)	0.000004 (0.000000)	0.000003 (0.000000)	0.000005 (0.000000)	0.000004 (0.000000)
alpha1	0.005482 (0.347708)	0.007907 (0.178516)	0.000000 (0.999999)	0.000000 (1.000000)	0.000000 (0.999992)	0.000000 (1.000000)
beta1	0.876664 (0.00000)	0.811451 (0.000000)	0.887942 (0.000000)	0.803632 (0.000000)	0.879190 (0.000000)	0.806809 (0.000000)
gamma1	0.148273 (0.00000)	0.250010 (0.000000)	0.151543 (0. 00000)	0.315522 (0.000000)	0.160613 (0.000000)	0.289259 (0.000000)
shape	-	-	9.459432 (0.000000)	5.216912 (0.000000)	1.555863 (0.000001)	1.244586 (0.000000)

Note: the highlighted results correspond to the coefficients that are not statistically significant.

In GJR-GARCH (1,1) model, the results are not so easy to interpret. In the case of the Russell 2000, some conditional mean estimates that are insignificant: the intercept, the first-order autoregressive coefficient, and the first-order moving average coefficient, under the normal distribution, and the first-order autoregressive and moving average coefficients, under the student-t and GED distributions. Regarding the conditional variance estimates, both the Russell 2000 and Russell 1000 have alpha1 as the only insignificant estimate.

The sum alpha1 and beta1 give values below 1 as desired. The beta1 estimates are higher to the Russell 2000, leading us to conclude that the shocks are more persistent to smaller firms, as already concluded by previous estimations. The gamma1 assumes positive values, leading to the confirmation of the leverage effect, which is in line with the EGARCH model results. As gamma1 is higher to the index that represents large firms, this model suggests that bigger firms tend to suffer more with bad news than smaller firms.

Table 4.16- APARCH (1,1) model estimates (the results in brackets represent the p-values)

	Normal Distribution		**Student-t Distribution**		**GED Distribution**	
	Russell 2000	**Russell 1000**	**Russell 2000**	**Russell 1000**	**Russell 2000**	**Russell 1000**
Conditional Mean Estimates						
mu	0.000327 (0.142210)	0.000754 (0.00000)	0.000750 (0.000000)	0.000494 (0.000008)	0.000717 (0.000000)	0.000578 (0.00000)
ar1	-0.461466 (0.000000)	0.168755 (0.00000)	-0.003055 (0.696957)	-1.941397 (0.00000)	0.284064 (0.000000)	0.173506 (0.00000)
ar2	-0.914098 (0.00000)	0.833437 (0.00000)	0.897745 (0.000000)	-0.942821 (0.00000)	0.625097 (0.000000)	0.701100 (0.00000)
ma1	0.437494 (0.000000)	-0.193381 (0.00000)	-0.030418 (0.000369)	1.924727 (0.00000)	-0.332999 (0.000000)	-0.219482 (0.00000)
ma2	0.904223 (0.00000)	-0.809005 (0.00000)	-0.901743 (0.000000)	0.925809 (0.00000)	-0.608938 (0.000000)	-0.679856 (0.00000)
Conditional Variance Estimates						
omega	0.000000 (0.920537)	0.000000 (0.945287)	0.000000 (0.852477)	0.000301 (0.043904)	0.000000 (0.891706)	0.000153 (0.274256)
alpha1	0.040227 (0.380881)	0.054776 (0.00004)	0.052254 (0.147586)	0.130785 (0.00000)	0.049627 (0.020137)	0.122870 (0.00000)
beta1	0.899470 (0.00000)	0.793466 (0.00000)	0.872788 (0.000000)	0.854776 (0.00000)	0.880203 (0.000000)	0.837823 (0.00000)
eta11	0.359249 (0.059534)	0.607892 (0.00000)	0.373325 (0.000337)	1.000000 (0.00000)	0.263108 (0.000616)	0.895991 (0.00001)

lambda	2.773365 (0.00000)	2.909401 (0.00000)	2.624991 (0.000000)	1.050871 (0.00000)	2.821246 (0.000000)	1.225206 (0.00000)
shape	-	-	8.811127 (0.000000)	5.453903 (0.00000)	1.503147 (0.000000)	1.249678 (0.00000)

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Regarding the APARCH (1,1) estimates, the Russell 1000 has all estimates statistically significant under the student-t distribution and, under normal distribution and GED distribution, only has insignificant omega. In terms of the Russell 2000, we will analyze one distribution at a time. Under the normal distribution, the conditional mean estimates for the intercept, omega, alpha1, and eta11 are not statistically significant. Under student-t the first-order autoregressive coefficient, omega, and alpha1 are insignificant. Finally, under GED distribution, only the omega fails at being statistically significant.

As in the models presented before, the persistence in volatility is measured by the sum of alpha1 and beta1. In accordance with previous results, the alpha1 presents lower values to the Russell 2000, indicating that the small firms have less short-term persistence. By contrary, the beta1 suggests longer term persistence to the small firms. This is in line with the previous models.

In the APARCH model the asymmetry coefficient is given by eta11, and it assumes higher values to the Russell 1000 index, confirming the conclusions of the EGARCH and GJR-GARCH model: the negative impact of news is confirmed and seems to be higher to larger firms. Regarding the power coefficient, represented by lambda, under the normal distribution, it is higher to the Russell 1000 under the normal distribution, while under the student-t and GED distributions it is higher to the Russell 2000. In these last two cases, the small firms seem to assume higher volatility.

Table 4.17- TGARCH (1,1) model estimates (the results in brackets represent the p-values)

	Normal Distribution		**Student-t Distribution**		**GED Distribution**	
	Russell 2000	**Russell 1000**	**Russell 2000**	**Russell 1000**	**Russell 2000**	**Russell 1000**
Conditional Mean Estimates						
mu	0.000263 (0.22252)	0.007494 (0.00000)	0.000454 (0.030285)	0.001706 (0.000000)	0.000599 (0.003453)	0.00187 (0.000000)
ar1	-1.959102 (0.00000)	1.353717 (0.00000)	0.292526 (0.000000)	0.917869 (0.000000)	-0.234454 (0.000000)	0.61092 (0.000000)
ar2	-0.960310 (0.00000)	-0.353188 (0.00000)	-0.980054 (0.000000)	0.082494 (0.000000)	0.618467 (0.000000)	0.38956 (0.000000)
ma1	1.945689 (0.00000)	-1.391022 (0.00000)	-0.292346 (0.000000)	-0.974553 (0.000000)	0.190578 (0.000000)	-0.67221 (0.000000)
ma2	0.946361 (0.00000)	0.403717 (0.00000)	0.994523 (0.082187)	0.024567 (0.000000)	-0.630666 (0.000000)	-0.32602 (0.000000)
Conditional Variance Estimates						
omega	0.000472 (0.00000)	0.000345 (0.00000)	0.000391 (0.000000)	0.000372 (0.000000)	0.000410 (0.000078)	0.00041 (0.000000)
alpha1	0.081067 (0.00000)	0.117793 (0.00000)	0.080901 (0.00000)	0.125563 (0.000000)	0.078975 (0.000000)	0.12151 (0.000000)
beta1	0.895713 (0.00000)	0.871132 (0.00000)	0.902553 (0.00000)	0.860403 (0.000000)	0.901045 (0.000000)	0.85717 (0.000000)
eta11	0.956151 (0.00000)	1.000000 (0.00000)	1.000000 (0.00000)	1.000000 (0.000000)	0.908135 (0.000000)	1.000000 (0.000000)
shape	-	-	9.809112 (0.000000)	5.552442 (0.000000)	1.562198 (0.000001)	1.25817 (0.000000)

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Lastly, we have the TGARCH (1,1) model. Regarding the significance of the estimates, this is one of the models with less problems of significance. The Russell 1000 has all estimates statistically significant, while in the Russell 2000 the conditional mean intercept is insignificant under the normal and student-t distributions, and the second-order moving average coefficient is insignificant under the student-t distribution. The results are in line with the previous models.

Table 4.18- Sign Bias test

	Model distribution	Model	p-value
Russell 2000	Normal Distribution	GARCH (1,1)	0.0002217
		EGARCH (1,1)	0.06763
		GJR-GARCH (1,1)	0.005673
		APARCH (1,1)	0.004072
		TGARCH (1,1)	0.11403
	Student-t Distribution	GARCH (1,1)	0.0001434
		EGARCH (1,1)	0.0495
		GJR-GARCH (1,1)	0.009114
		APARCH (1,1)	0.001946
		TGARCH (1,1)	0.1434
	GED	GARCH (1,1)	0.0002409
		EGARCH (1,1)	0.06672
		GJR-GARCH (1,1)	0.005729
		APARCH (1,1)	0.001937
		TGARCH (1,1)	0.05942
Russell 1000	Normal Distribution	GARCH (1,1)	0.0004227
		EGARCH (1,1)	0.1524
		GJR-GARCH (1,1)	0.003306
		APARCH (1,1)	0.004515
		TGARCH (1,1)	0.06805
	Student-t Distribution	GARCH (1,1)	0.0004227
		EGARCH (1,1)	0.2125
		GJR-GARCH (1,1)	0.02862
		APARCH (1,1)	0.15939
		TGARCH (1,1)	0.13199
	GED	GARCH (1,1)	0.0003412
		EGARCH (1,1)	0.1884
		GJR-GARCH (1,1)	0.03805
		APARCH (1,1)	0.003800
		TGARCH (1,1)	0.03280

By analyzing the sign bias test in Table 4.18, we can see that in the GARCH model, the p-value is lower than the significance level, meaning the leverage effect was not considered in the estimation of the model. On contrary, in the asymmetric models, although this is not so obvious in some of them, the leverage effect is already contemplated. As the p-value is above the significance level, the null hypothesis is not rejected, and the leverage effect was well captured by the models.

Table 4.19- Information Criteria Measures

			AIC	BIC
Russell 2000	Normal Distribution	GARCH (1,1)	-6.1951	-6.1728
		EGARCH (1,1)	-6.2263	-6.2013
		GJR-GARCH (1,1)	-6.2184	-6.1934
		APARCH (1,1)	-6.2084	-6.1806
		TGARCH (1,1)	-6.2291	-6.2040
	Student-t Distribution	GARCH (1,1)	-6.2169	-6.1919
		EGARCH (1,1)	-6.2432	-6.2154
		GJR-GARCH (1,1)	-6.2352	-6.2074
		APARCH (1,1)	-6.2287	-6.1981
		TGARCH (1,1)	-6.2457	-6.2178
	GED	GARCH (1,1)	-6.2173	-6.1922
		EGARCH (1,1)	-6.2406	-6.2127
		GJR-GARCH (1,1)	-6.2336	-6.2058
		APARCH (1,1)	-6.2254	-6.1947
		TGARCH (1,1)	-6.2394	-6.2115
Russell 1000	Normal Distribution	GARCH (1,1)	-6.8409	-6.8186
		EGARCH (1,1)	-6.8885	-6.8634
		GJR-GARCH (1,1)	-6.8781	-6.8531
		APARCH (1,1)	-6.8666	-6.8388
		TGARCH (1,1)	-6.8886	-6.8636
	Student-t Distribution	GARCH (1,1)	-6.8409	-6.8186
		EGARCH (1,1)	-6.9456	-6.9177
		GJR-GARCH (1,1)	-6.9402	-6.9124
		APARCH (1,1)	-6.9470	-6.9164
		TGARCH (1,1)	-6.9525	-6.9247
	GED	GARCH (1,1)	-6.9103	-6.8853
		EGARCH (1,1)	-6.9467	-6.9189
		GJR-GARCH (1,1)	-6.9412	-6.9134
		APARCH (1,1)	-6.9448	-6.9142
		TGARCH (1,1)	-6.9501	-6.9223

Note: the results highlighted in grey correspond to the lowest information criteria values.

Concerning the worst model, it ended up being the GARCH model, as we were already excepting, given its characteristics. Within the asymmetric models, we can conclude by the analysis of the information criteria that the best model to estimate the Russell 2000 and the Russell 1000 is the TGARCH (1,1) under the student-t distribution.

Table 4.20- ARCH LM Tests applied to the Standardized Squared Residuals

		Russell 2000	**Russell 1000**
ARCH LM Test	ARCH Lag [3]	0.2099	0.14059
	ARCH Lag [5]	0.3555	0.37458
	ARCH Lag [7]	0.5245	0.01174

Based on the ARCH LM test applied to the squared residuals, as the p-value is higher than 5%, we confirm there is no ARCH effect anymore.

Table 4.21- Ljung-Box Test applied to the Standardized and Standardized Squared Residuals

		Russell 2000	**Russell 1000**
Ljung-Box Test (Standardized Residuals)	Lag [1]	0.07858	0.5723
	Lag [11]	0.02340	0.9996
	Lag [19]	0.07269	0.5340
Ljung-Box Test (Standardized Squared Residuals)	Lag [1]	0.6059	0.88469
	Lag [5]	0.3765	0.73582
	Lag [9]	0.5256	0.09983

The results of the Ljung-Box test applied to the standardized residuals point to the lack of autocorrelation of the residuals, as the p-value is higher than the significance level 5%, meaning that the models fit well the data and that there is no autocorrelation anymore, revealing the white noise behavior of the residuals (in terms of autocorrelation). Regarding the Ljung-Box test applied to the standardized squared residuals, the test result confirms the one obtained through the ARCH LM test in Table 4.20.

In Figure 4.20 we have the news impact curve, obtained through the function "newsimpact". We can see that the greater the shock, the greater is the impact on volatility. In the worst case scenario ($\varepsilon_{t-1} = -0.3$), the index of large companies reaches a higher level of volatility when compared to small companies. This is in line with the conclusion we took in the EGARCH model, GJR-GARCH model, APARCH model, and TGARCH model.

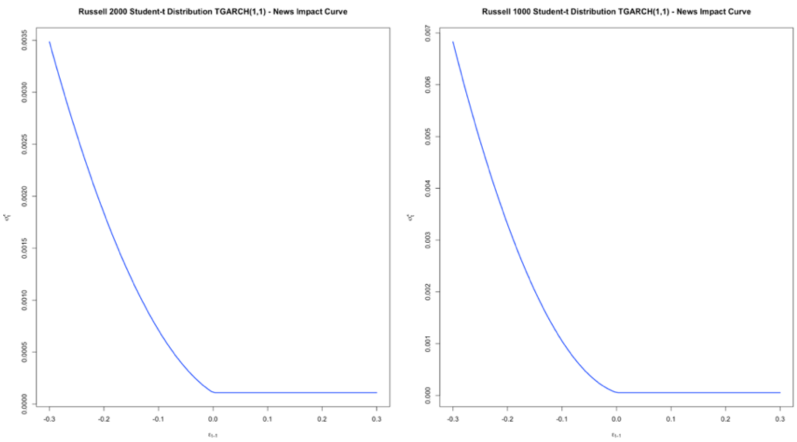

Figure 4.20- News Impact Curve of the TGARCH (1,1) Model with Student-t Distribution

4.4.2 Out-of-sample analysis

To forecast the variance of our series, we will rely on the "ugarchforecast" function of the rugarch package. With this function, we can do a normal forecast or a rolling window forecast, by establishing the condition n.roll > 0. In our study we opted by a rolling window forecast and defined the function as follows: (1) in the "fitORspec" we considered the model specified in the in-sample analysis, (2) the forecast horizon was defined as n.ahead=1, and (3) the in-sample part of the sample as considered as the data. Regarding the size of the rolling window, we follow the procedure of Costa (2017) who, for a sample of 30 years of data, considered three rolling window sizes (500, 1000, and 2000), and the loss functions to see which produces the most desirable results. As our sample is one-third of the one of him, the length of the out-of-sample part obviously needs to be much smaller. So, considering a similar proportion of him, we consider the rolling window sizes equal to 150, 250, and 500.

Table 4.22- Out-of-sample Loss Functions (rolling window size = 150)

			MAE	**MSE**	**RMSE**	**MPE**
Russell 2000	Normal Distribution	GARCH (1,1)	0.00974801	9.933115e-05	0.009966501	-7482.911
		EGARCH (1,1)	0.009569345	9.901964e-05	0.009950861	-7345.762
		GJR-GARCH (1,1)	0.009968744	0.0001079851	0.01039159	-7652.355
		APARCH (1,1)	0.009935126	0.000106549	0.01032226	-7626.548
		TGARCH (1,1)	0.009628144	0.0001002336	0.01001167	-7390.898
	Student-t Distribution	GARCH (1,1)	0.00969853	9.856645e-05	0.009928064	-7444.929
		EGARCH (1,1)	0.009470779	9.81173e-05	0.009905417	-7270.099
		GJR-GARCH (1,1)	0.009963287	0.0001090672	0.01044353	-7648.166
		APARCH (1,1)	0.01002689	0.0001097942	0.01047827	-7696.99
		TGARCH (1,1)	0.009601962	0.0001004446	0.01002221	-7370.8
	GED Distribution	GARCH (1,1)	0.00972828	9.928793e-05	0.009964333	-7467.766
		EGARCH (1,1)	0.009536755	9.881266e-05	0.009940456	-7320.745
		GJR-GARCH (1,1)	0.009925408	0.0001077877	0.01038209	-7619.088
		APARCH (1,1)	0.009966427	0.0001077205	0.01037885	-7650.576
		TGARCH (1,1)	0.009617487	0.0001001669	0.01000834	-7382.718
Russell 1000	Normal Distribution	GARCH (1,1)	0.008575753	8.63918e-05	0.009294719	-11546.59
		EGARCH (1,1)	0.008384099	8.6349e-05	0.009292416	-11288.54
		GJR-GARCH (1,1)	0.008803	9.883435e-05	0.009941547	-11852.56
		APARCH (1,1)	0.008759997	0.0001010622	0.01005297	-11794.66
		TGARCH (1,1)	0.008515273	9.150533e-05	0.009565842	-11465.16
	Student-t Distribution	GARCH (1,1)	0.008575753	8.63918e-05	0.009294719	-11546.59
		EGARCH (1,1)	0.008594657	9.405119e-05	0.009697999	-11572.04
		GJR-GARCH (1,1)	0.009132695	0.0001117689	0.01057208	-12296.47
		APARCH (1,1)	0.008809985	9.903203e-05	0.009951484	-12307.2
		TGARCH (1,1)	0.008341287	8.940135e-05	0.009455229	-11230.9
	GED Distribution	GARCH (1,1)	0.008638807	8.888032e-05	0.009427636	-11631.49
		EGARCH (1,1)	0.00845968	8.998204e-05	0.009485886	-11390.31
		GJR-GARCH (1,1)	0.008949544	0.0001058005	0.01028594	-12502.16
		APARCH (1,1)	0.008697265	9.614114e-05	0.009805159	-11710.19
		TGARCH (1,1)	0.008221612	8.565862e-05	0.009255194	-11069.77

Note: the highlighted results correspond to the smallest loss function values.

Table 4.23- Out-of-sample Loss Functions (rolling window size = 250)

			MAE	**MSE**	**RMSE**	**MPE**
Russell 2000	Normal Distribution	GARCH (1,1)	0.009096771	8.624781e-05	0.00928697	-6982.997
		EGARCH (1,1)	0.008842096	8.470834e-05	0.009203713	-6787.5
		GJR-GARCH (1,1)	0.009238573	9.188248e-05	0.009585535	-7091.85
		APARCH (1,1)	0.009158076	8.981417e-05	0.009477034	-7030.058
		TGARCH (1,1)	0.008971994	8.680921e-05	0.009317146	-6887.215
	Student-t Distribution	GARCH (1,1)	0.00898191	8.441644e-05	0.009187842	-6894.826
		EGARCH (1,1)	0.008712503	8.329084e-05	0.009126382	-6688.021
		GJR-GARCH (1,1)	0.009175791	9.17018e-05	0.009576106	-7043.656
		APARCH (1,1)	0.00926296	9.270876e-05	0.009628539	-7110.57
		TGARCH (1,1)	0.008892501	8.611928e-05	0.009280047	-6826.193
	GED Distribution	GARCH (1,1)	0.009045924	8.561285e-05	0.009252721	-6943.966
		EGARCH (1,1)	0.008794183	8.422679e-05	0.009177515	-6750.721
		GJR-GARCH (1,1)	0.00917047	9.118469e-05	0.009549068	-7039.572
		APARCH (1,1)	0.00920791	9.104699e-05	0.009541855	-7068.312
		TGARCH (1,1)	0.008948853	8.661189e-05	0.009306551	-6869.451
Russell 1000	Normal Distribution	GARCH (1,1)	0.007315047	6.39782e-05	0.007998637	-9849.145
		EGARCH (1,1)	0.007097424	6.321803e-05	0.007950977	-9556.133
		GJR-GARCH (1,1)	0.007538531	7.261942e-05	0.008521703	-10150.05
		APARCH (1,1)	0.007518343	7.400673e-05	0.008602717	-10122.87
		TGARCH (1,1)	0.007045891	6.489713e-05	0.008055875	-9486.747
	Student-t Distribution	GARCH (1,1)	0.007315047	6.39782e-05	0.007998637	-9849.145
		EGARCH (1,1)	0.007201841	6.782906e-05	0.00823584	-9696.722
		GJR-GARCH (1,1)	0.007698939	8.026059e-05	0.008958827	-10366.03
		APARCH (1,1)	0.007344141	7.115089e-05	0.008435099	-9888.319
		TGARCH (1,1)	0.00690591	6.337316e-05	0.007960726	-9298.274
	GED Distribution	GARCH (1,1)	0.007281629	6.476662e-05	0.008047771	-9804.15
		EGARCH (1,1)	0.007123838	6.534296e-05	0.008083499	-9591.696
		GJR-GARCH (1,1)	0.007586069	7.649119e-05	0.008745924	-10214.05
		APARCH (1,1)	0.007329075	6.952797e-05	0.008338343	-9868.033
		TGARCH (1,1)	0.006859419	6.130593e-05	0.007829811	-9235.678

Note: the highlighted results correspond to the smallest loss function values.

Table 4.24- Out-of-sample Loss Functions (rolling window size = 500)

			MAE	MSE	RMSE	MPE
Russell 2000	Normal Distribution	GARCH (1,1)	0.009362832	9.061018e-05	0.009518938	-7187.236
		EGARCH (1,1)	0.009187828	8.886905e-05	0.009427038	-7052.896
		GJR-GARCH (1,1)	0.009377556	9.23935e-05	0.009612154	-7198.538
		APARCH (1,1)	0.009411219	9.293506e-05	0.009640283	-7224.379
		TGARCH (1,1)	0.009195052	8.894421e-05	0.009431024	-7058.442
	Student-t Distribution	GARCH (1,1)	0.009290694	8.93286e-05	0.009451381	-7131.86
		EGARCH (1,1)	0.009098004	8.779129e-05	0.0093697	-6983.944
		GJR-GARCH (1,1)	0.009338915	9.231679e-05	0.009608163	-7168.876
		APARCH (1,1)	0.00942651	9.382249e-05	0.009686201	-7236.117
		TGARCH (1,1)	0.009122751	8.803564e-05	0.009382731	-7002.941
	GED Distribution	GARCH (1,1)	0.00931784	8.987794e-05	0.009480398	-7152.698
		EGARCH (1,1)	0.009154241	8.849423e-05	0.009407137	-7027.114
		GJR-GARCH (1,1)	0.009330517	9.191462e-05	0.009587211	-7162.429
		APARCH (1,1)	0.009395108	9.280195e-05	0.009633377	-7212.012
		TGARCH (1,1)	0.009203291	8.923079e-05	0.009446205	-7064.766
Russell 1000	Normal Distribution	GARCH (1,1)	0.006711045	5.156324e-05	0.007180755	-9035.903
		EGARCH (1,1)	0.006616419	5.166629e-05	0.007187927	-8908.497
		GJR-GARCH (1,1)	0.006837636	5.63506e-05	0.007506704	-9206.349
		APARCH (1,1)	0.006830005	5.741847e-05	0.007577498	-9196.073
		TGARCH (1,1)	0.006436601	5.042796e-05	0.007101264	-8666.386
	Student-t Distribution	GARCH (1,1)	0.006711045	5.156324e-05	0.007180755	-9035.903
		EGARCH (1,1)	0.006702999	5.475009e-05	0.007399331	-9025.07
		GJR-GARCH (1,1)	0.006914903	6.062344e-05	0.007786106	-9310.383
		APARCH (1,1)	0.006670217	5.488928e-05	0.00740873	-8980.932
		TGARCH (1,1)	0.006296262	4.902417e-05	0.007001726	-8477.429
	GED Distribution	GARCH (1,1)	0.006646178	5.149369e-05	0.00717591	-8948.565
		EGARCH (1,1)	0.006634776	5.302413e-05	0.007281767	-8933.213
		GJR-GARCH (1,1)	0.006834864	5.824316e-05	0.007631721	-9202.616
		APARCH (1,1)	0.006683731	5.421477e-05	0.007363068	-8999.127
		TGARCH (1,1)	0.00625536	4.767109e-05	0.006904426	-8422.358

Note: the highlighted results correspond to the smallest loss function values.

In Tables 4.25, 4.26, and 4.27 we present the loss functions results obtained for the three rolling window sizes. Based on these tables we can see that the results were not very sensitive to the size of the rolling window, as the loss function point to the same conclusion in all three cases. Relying on MAE, MSE, and RMSE, the best model to fit both series are: (1) Russell 2000: EGARCH (1,1) under student-t distribution with rolling window size equal to 250, (2) Russell 1000: TGARCH (1,1) under GED distribution with rolling window size equal to 500. On the other hand, if we consider the MPE, other models are preferred: (1) Russell 2000: APARCH (1,1) under student-t distribution, (2) Russell 1000: GJR-GARCH (1,1) under student-t distribution. We opted by considering the models chosen by the majority of the loss functions. As we can see in the figures displayed below, the unconditional variance of the Russell 2000 shows a growing trend (Figure 4.21), while the rolling variance shows an high oscillation through time, with a tendency of decrease at the end of our forecast series (Figure 4.22):

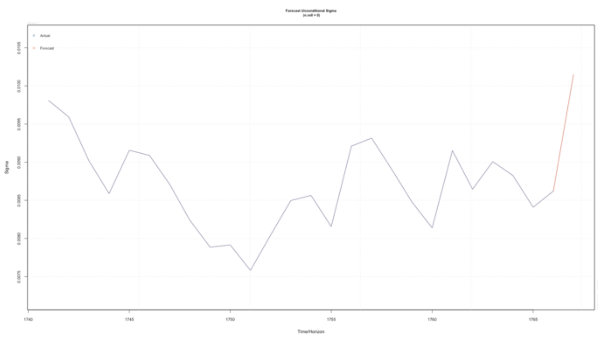

Figure 4.21- Forecast Unconditional Sigma (n.roll=0) for Russell 2000

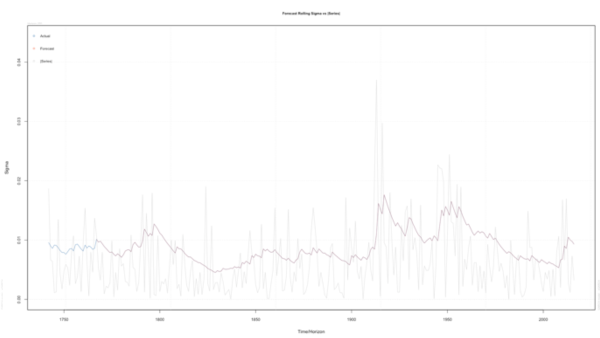

Figure 4.22- Forecast Rolling Sigma vs |Series| for Russell 2000

Contrary to the Russell 2000, the Russell 1000 unconditional sigma starts decreasing at the end of the forecast series (Figure 4.23), and the rolling sigma, although more stable through a long period, records a slightly higher variance at almost the end of forecast (Figure 4.24):

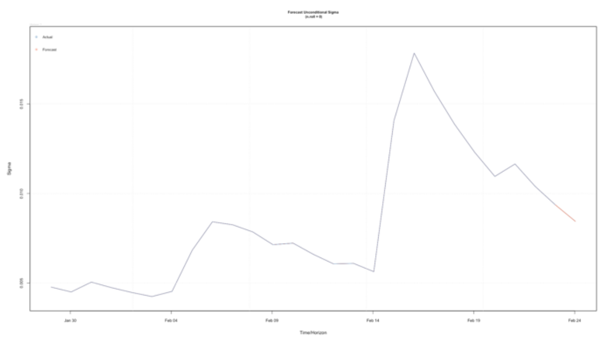

Figure 4.23- Forecast Unconditional Sigma (n.roll=0) for Russell 1000

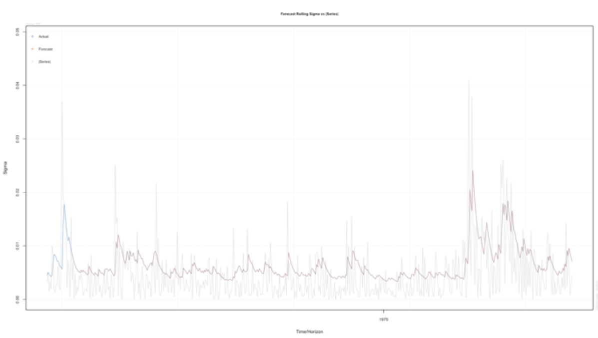

Figure 4.24- Forecast Rolling Sigma vs |Series| for Russell 1000

4.5 Multivariate Conditional Variance Models

4.5.1 VAR Model

4.5.1.1 VAR Model Estimation

In order to confirm if two series Granger cause each other we can run the granger causality test in R program by performing the "causality" function, with the following options:

1. Russell 2000 Granger cause Russell 1000: x and y are the Russell 2000 and Russell 1000 time series returns, respectively;
2. Russell 1000 Granger cause Russell 2000: x and y are the Russell 1000 and Russell 2000 time series returns, respectively.

Considering that the null hypotheses of the test is defined as non-causation we conclude that, as the p-values are less than the significance level 5% (Annex H), we can conclude that both null hypotheses are rejected, leading to the conclusion that the Russell 2000 does Granger cause Russell 1000, and vice-versa.

As we have seen before, the stationarity condition of both series was already checked before estimating the univariate conditional variance models. Therefore, we can proceed with the estimation of the VAR model.

To choose properly the order of the VAR model, we opted by running the "VARselect" function, where we considered a time series constituted by the two indexes daily returns, and the type of deterministic regressors was kept on the default setting ("const"). Based on the HQ and SC results, the best order is 8 and 3, respectively. We will rely on the result given by the SC, as suggested by Hyndman & Athanasopoulos (2018), and consider estimate a more parsimonious model.

4.5.1.2 Diagnostic Testing

After estimating the multivariate model, we performed the multivariate ARCH test and concluded that, as the p-value is lower than the significance level 5%, there exists heteroscedasticity (Annex I). To check the autocorrelation, we used the portmanteau tests and the Breusch-Godfrey LM test. All three tests point to the same conclusion: as the p-value is below the significance level 5%, we should reject the null hypotheses and state that there exists serial correlation (Table 4.25).

Table 4.25- Multivariate Autocorrelation Tests

	Portmanteau Test (asymptotic)	**Portmanteau Test (adjusted)**	**Breusch-Godfrey LM test**
χ^2	190.16	190.84	85.981
p-value	< 2.2e-16	< 2.2e-16	3.707e-10

Regarding the multivariate normality tests, we can see in Annex J that the normality distribution of the residuals is not confirmed.

In Figure 4.25 we have the CUSUM test, which is usually used to do a stability test. Based on this plot, as the red line was never exceeded, we can assume that there are no structural breaks.

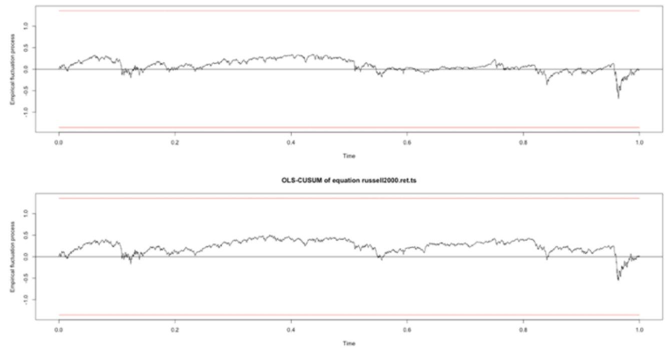

Figure 4.25- Stability test plot

4.5.1.3 Policy Simulations

Starting with the impulse response analysis, we see in the figures below that a shock in Russell 1000 has a big impact on the returns of the index of small firms (Figure 4.26). On the contrary, the shocks in Russell 2000 do not seem to affect that much the index Russell 1000 returns (Figure 4.27). We can also notice that in Figure 4.27 the responses become insignificant after period 6, while in Figure 4.26 this happens earlier in period 4:

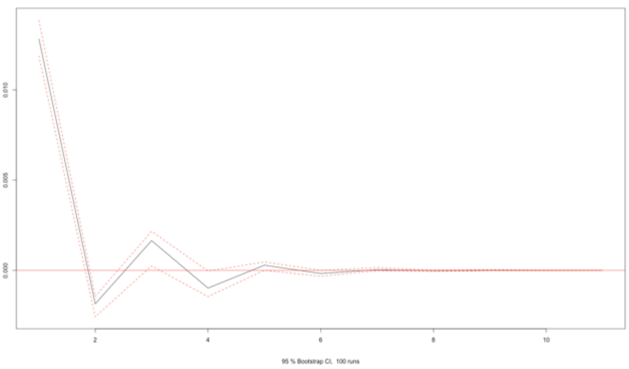

Figure 4.26- Russell 2000 Impulse Response to a shock in Russell 1000

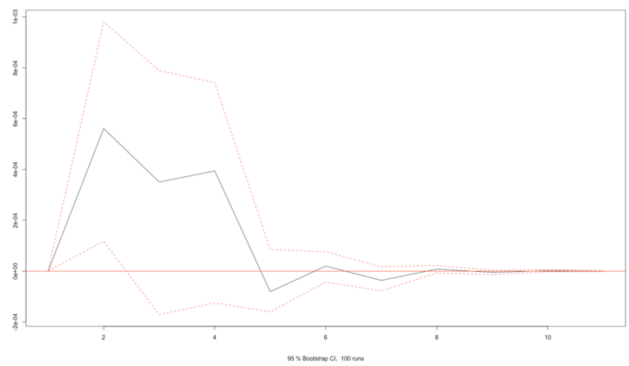

Figure 4.27- Russell 1000 Impulse Response to a shock in Russell 2000

Next, we present the Forecast Error Variance Decomposition (FEVD), which allows us to check if the error variance of one variable contributes to explain the error variance of other variables. Based on the FEVD presented in Table 4.26, we can conclude that the Russell 1000 contributes more to the variance of the error of the Russell 2000, than the Russell 2000 does to the Russell 1000. We can see that in 5 days, the Russell 2000 returns contribute only 0.4% to the Russell 1000 returns while the Russell 1000 returns contribute 84.78849% to the Russell 2000 returns.

Table 4.26- Table of Forecast Error Variance Decomposition Summary

$russell1000.ret.ts		
	russell1000.ret.ts	russell2000.ret.ts
[1,]	1.0000000	0.000000000
[2,]	0.9973890	0.002610974
[3,]	0.9964170	0.003582969
[4,]	0.9951585	0.004841535
[5,]	0.9951062	0.004893779
$russell2000.ret.ts		
	russell1000.ret.ts	russell2000.ret.ts
[1,]	0.8437904	0.1562096
[2,]	0.8458409	0.1541591
[3,]	0.8472378	0.1527622
[4,]	0.8478528	0.1521472
[5,]	0.8478849	0.1521151

This can be confirmed by the analysis of Figure 4.28, where we can observe that the variation of Russell 1000 returns contribute more to Russell 2000 returns variation than vice versa:

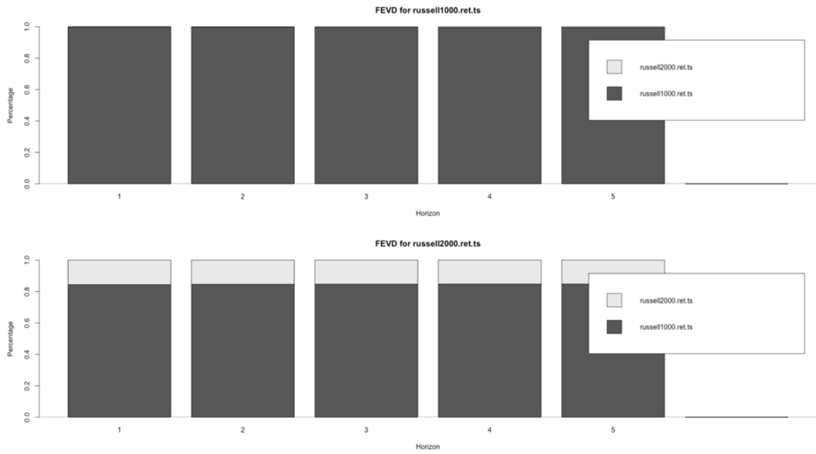

Figure 4.28- Forecast Error Variance Decomposition Plot

So, according to both impulse response and forecast error variance decomposition analysis, we can conclude that there are transmissions of volatility from large to small firms, contributing to explain some of its behavior, while the small firms do not contribute as much to explain the behavior of large firms volatility.

4.5.2 DCC Model

After performing the test proposed by Engle & Sheppard (2001) to decide between constant and dynamic correlations, we obtained a p-value (0.03418031) lower than the 5% significance level, leading us to reject the null hypotheses. Therefore, we are in the presence of non-constant conditional correlations, which demands a more adequate model as the DCC model. For this reason, we do not present the estimation results of CCC model.

We started by estimating the asymmetric model GJR-GARCH (1,1) as Billio, Caporin, & Gobbo (2006) did, using the same R function used in the Univariate Conditional Variance Models section. In order to do the specification of the model we used the function “dccspec”, considering two different distributions (normal and student-t distributions). For the estimation we considered the “dccfit” function.

Table 4.27- DCC-GARCH model results (out of sample = 500)

		Normal Distribution	Student-t Distribution
GJR-GARCH (1,1)			
Russell 1000	omega	0.000003 (0.000001)	0.000003 (0.151619)
	alpha1	0.012058 (0.310202)	0.000002 (0.999949)
	beta1	0.817249 (0.000000)	0.814906 (0.000000)
	gamma1	0.226346 (0.000000)	0.294585 (0.000000)
	shape	-	5.292852 (0.000000)
Russell 2000	omega	0.000005 (0.000000)	0.000004 (0.024931)
	alpha1	0.005284 (0.514766)	0.000004 (0.999664)
	beta1	0.875671 (0.00000)	0.889367 (0.00000)
	gamma1	0.142289 (0.00001)	0.147700 (0.00001)
	shape	-	8.852236 (0.00000)
DCC (1,1)			
	dcca1	0.064790 (0.00000)	0.074245 (0.000000)
	dccb1	0.896467 (0.00000)	0.884047 (0.00000)
	mshape	-	7.489999 (0.000000)
Information Criteria			
AIC		-14.786	-14.843
BIC		-14.739	-14.788

Note: the highlighted results correspond to the coefficients that are not statistically significant.

As we did in the univariate case, we considered three rolling windows but, in order to do not waste space on this report, we only present the DCC model with a rolling window equal to 500, as it is the one that shows lower information criteria values (the remaining can be assessed in the Annex K and Annex L). Additional to this we also estimated the aDCC and FDCC models and, for the same reason, they are presented from Annex M to Annex R. Please note that due to a program constrain, we could not obtain any results for the FDCC model with student-t distribution.

In terms of the estimates, we have some insignificant estimates for the coefficients: (1) alpha1 to both Russell 2000 and 1000, under normal and student-t distribution, and (2) Russell 1000 omega under student-t distribution. As we already pointed earlier, this is not a problem. Regarding the DCC joint estimates, both dcca1 and dccb1 are statistically significant, confirming the time-varying characteristic of the conditional correlations. This supports the results obtained through the estimation of the Engle & Sheppard (2001) test, confirming that the DCC is a better choice, when compared with the CCC. As expected, the dcca1 presents values close to zero, while dccb1 assumes values of almost one. We also confirm that the model is positive definite, as both dcca1 and dcca1 assume positive values and the sum of the dcca1 and dccb1 produces a value below one, as stated by Engle & Sheppard (2001).

Considering the student-t distribution, as it is expected to produce better results than the normal distribution since its lower information criteria values, we obtain the following graphs:

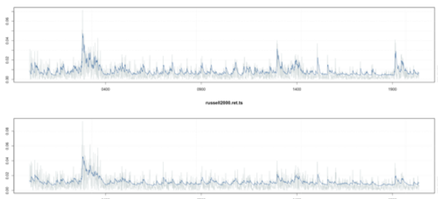

Figure 4.29- DCC Conditional Sigma vs |returns|

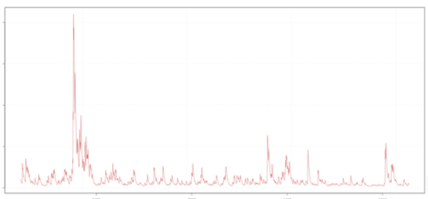

Figure 4.30- DCC Conditional Covariance

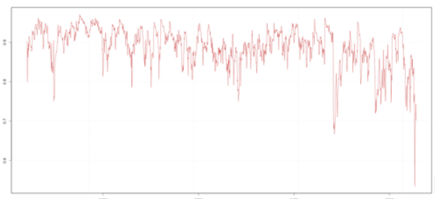

Figure 4.31- DCC Conditional Correlation

In Figure 4.29, Figure 4.30, and Figure 4.31 we can see the behavior of DCC conditional variance, DCC conditional covariance, and DCC conditional correlation, respectively. In terms of the conditional variance and conditional covariance, they follow a similar path and tend to record an increase in periods of crisis. Regarding the conditional correlation, they show a quite unstable behavior through time, presenting sharper rises on more volatility periods, supporting the time-varying conditional correlation characteristic already seen previously. By analyzing these three figures, we can conclude the existence of volatility spillovers, as the conditional correlation tends to rise in periods of crisis, implying that both indexes affect each other.

Moving to the forecasting part, we opted by the same rolling forecast as we did in the Univariate Conditional Variance Models section:

Figure 4.32- DCC Sigma Rolling Forecast for Russell 1000 and Russell 2000

Figure 4.33- DCC Covariance Rolling Forecast for Russell 1000 and Russell 2000

Both rolling variance (Figure 4.32) and rolling covariance (Figure 4.33) show a similar tendency, recording a stabilization that seems to maintain in a long term perspective, with an high peak at the end of our forecast series.

5 Conclusion

The firm size is the center of many studies nowadays, being the impact of the firm size on the returns a subject already deeply investigated, pointing to the well-known size effect. Banz (1981) is the first to question whether investors were indeed rewarded by an increased return when exposed to more risky assets. Besides this extensive literature, not many researchers focused on the volatility of such returns, which led to an open door to our study. Chelley-Steeley & Steeley (1995, 1996) belong to the small group of authors who focused on understanding this relationship, and the possibility of existing volatility spillovers between the two different types of firms (small and large firms).

To conclude which type of firms are more affected by the shocks, we based our study on the univariate models (GARCH, EGARCH, GJR-GARCH, APARCH, e TGARCH), while to study the volatility transmissions, the multivariate models seemed more adequate (VAR and DCC). Besides this, we also divided our sample into two parts to know which model was the most accurate to obtain our conclusions, as a good in-sample model may not be the best to forecast volatility. This was exactly what happened in our study. Based on the in-sample period, the best model for both indexes was the TGARCH (1,1) under the student-t distribution, while in the out-of-sample analysis, the results differed between indexes. To the Russell 2000, the most accurate model was the EGARCH (1,1) under student-t distribution with rolling window size equal to 250, while to the Russell 1000 the TGARCH (1,1) under GED distribution with rolling window size equal to 500 stood out. Despite the choice of the model being different between the two indexes, all univariate models led to similar conclusions.

In line with Chelley-Steeley & Steeley (1995, 1996), small firms are more affected in the short-run, revealing higher persistence of shocks in this type of firm, while large firms are more affected in the long-run perspective. Nevertheless, contrary to them, the leverage effect was confirmed to be higher to larger firms, meaning that large firms tend to be more sensitive to negative shocks. Based on the APARCH model we can also conclude that there is evidence to affirm that small firm's returns tend to be more volatile, which is in accordance with the studies mentioned in literature review (Baskin, 1989; Habib, Kiani, & Khan, 2012; Hussainey, Mgbame, & Chijoke-Mgbame, 2011; Nazir, Nawaz, Anwar, & Ahmed, 2010).

In terms of the volatility spillovers across small and large firms, we opted by considering the multivariate models VAR and DCC. Based on these models, we confirmed the existence of volatility transmissions between small and large firms. By estimating the VAR model, we can understand by analyzing the impulse response plot and FEVD that the volatility of one index has some impact on the other. More specifically, we found that the index representing smaller firms tend to react more strongly to shocks in the Russell 1000 index, than the opposite. Unfortunately, this model is not enough as our primary focus is on the conditional correlations between both indexes. Considering this, we used the DCC, as it ends up being a better choice as the conditional correlations do not show to be constant over time. In light of the DCC, we confirmed the increase of conditional correlation between Russell 2000 and Russell 1000 in periods of higher volatility, and an unstable conditional variance and covariance, depending on if we are looking at the periods of crisis or not. This meets the conclusion of Chelley-Steeley & Steeley (1996).

Bibliography

Agarwal, N. C. (1979). Nature of Size—Structure Relationship: Some Further Evidence. In *Human Relations* (Vol. 32). https://doi.org/10.1177/001872677903200601

Al-Khazali, O. M., & Zoubi, T. A. (2005). Empirical testing of different alternative proxy measures for firm size. *Journal of Applied Business Research*, *21*(3), 79–90. https://doi.org/10.19030/jabr.v21i3.1471

Amihud, Y., & Mendelson, H. (1987). Trading Mechanisms and Stock Returns: An Empirical Investigation. *The Journal of Finance*, *42*(3), 533–553. https://doi.org/10.2307/2328369

Anderson, O. D. (1977). The Box-Jenkins approach to time series analysis. *RAIRO. RECHERCHE OPÉRATIONNELLE*, *11*(1), 3–29.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*, 3–18.

Baskin, J. (1989). Dividend policy and the volatility of common stocks. *Journal of Portfolio Management*, *15*(3), 19–25.

Beaver, W. H. (1968). The Information Content of Annual Earnings Announcements. *Journal of Accounting Research*, *6*, 67–92. https://doi.org/10.2307/2490070

Beckers, S. (1980). The Constant Elasticity of Variance Model and Its Implications For Option Pricing. *The Journal of Finance*, *35*(3), 661–673. https://doi.org/10.2307/2327490

Bera, A. K., & Higgins, M. L. (1993). Arch Models: Properties, Estimation and Testing. *Journal of Economic Surveys*, *7*(4), 305–366. https://doi.org/10.1111/j.1467-6419.1993.tb00170.x

Berk, J. B. (1995). A Critique of Size-Related Anomalies. *The Review of Financial Studies*, *8*(2), 275–286. https://doi.org/10.1093/rfs/8.2.275

Billio, M., Caporin, M., & Gobbo, M. (2006). Flexible Dynamic Conditional Correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters*, *2*(2), 123–130. https://doi.org/10.1080/17446540500428843

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meeting of Business and Economics Statistics Section of the American Statistical Association*, 27, 399 – 418.

Bollerslev, T. (1986). GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Bollerslev, T. (1990). Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model. *The Review of Economics and Statistics*, *72*(3), 498–505. https://doi.org/10.2307/2109358

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance. A review of the theory and empirical evidence. *Journal of Econometrics*, *52*(1–2), 5–59. https://doi.org/https://doi.org/10.1016/0304-4076(92)90064-X

Bollerslev, T., Engle, R. F., & Nelson, D. B. B. T.-H. of E. (1994). *Chapter 49 Arch models*. https://doi.org/https://doi.org/10.1016/S1573-4412(05)80018-2

Bonaccorsi, A. (1992). On the Relationship between Firm Size and Export Intensity. *Journal of International Business Studies*, *23*(4), 605–635.

Box, G. E.P., & Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, *65*(332), 1509–1526. https://doi.org/10.2307/2284333

Box, George E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis : forecasting and control* (Fifth edit). New Jersey: John Wiley & Sons, Inc.

Bullock, N., & Dunkley, E. (2018, December 31). *Stock market slide in 2018 leaves investors bruised and wary*. Retrieved on August 30, 2020, from https://www.ft.com/content/73d3dd26-0ce0-11e9-acdc-4d9976f1533b

Calof, J. L. (1994). The Relationship Between Firm Size and Export Behavior Revisited. *Journal of International Business Studies*, *25*(2), 367–387.

Campbell, J. Y., & Hentschel, L. (1992). No news is good news. An asymmetric model of changing volatility in stock returns. In *Journal of Financial Economics* (Vol. 31). https://doi.org/10.1016/0304-405X(92)90037-X

Cappiello, L., Engle, R. F., & Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, *4*(4), 537–572. https://doi.org/https://doi.org/10.1093/jjfinec/nbl005

Charitou, A., Clubb, C., & Andreou, A. (2001). The Effect of Earnings Permanence, Growth, and Firm Size on the Usefulness of Cash Flows and Earnings in Explaining Security Returns: Empirical Evidence for the UK. *Journal of Business Finance and Accounting*, *28*(5–6), 563–594. https://doi.org/10.1111/1468-5957.00385

Checking normality for parametric tests in R. Retrieved on August 2, 2020, from https://www.sheffield.ac.uk/polopoly_fs/1.885202!/file/95_Normality_Check.pdf

Chelley-Steeley, P. L., & Steeley, J. M. (1995). Conditional volatility and firm size: An empirical analysis of UK equity portfolios. *Applied Financial Economics*, *5*(6), 433–440. https://doi.org/10.1080/758538603

Chelley-Steeley, P. L., & Steeley, J. M. (1996). VOLATILITY, LEVERAGE AND FIRM SIZE: THE U.K. EVIDENCE. *The Manchester School Supplement*, *64*(S1), 83–103. https://doi.org/10.1111/j.1467-9957.1996.tb01456.x

Cheung, Y., & Ng, L. K. (1992). Stock Price Dynamics and Firm Size: An Empirical Investigation. *The Journal of Finance*, *47*(5), 1985–1997.

https://doi.org/10.2307/2329006

Child, J. (1973). Predicting and Understanding Organization Structure. *Administrative Science Quarterly*, *18*(2), 168–185. https://doi.org/10.2307/2392061

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, *10*(4), 407-432. https://doi.org/10.1016/0304-405X(82)90018-6

Conrad, J., Gultekin, M. N., & Kaul, G. (1991). Asymmetric Predictability of Conditional Variances. *The Review of Financial Studies*, *4*(4), 597–622. https://doi.org/https://doi.org/10.1093/rfs/4.4.597

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. In *Quantitative Finance* (Vol. 1). https://doi.org/10.1080/713665670

Coronavirus creates a true 'bear market' for Wall Street. (2020, March 12). Retrieved on August 30, 2020, from https://www.presstelegram.com/2020/03/12/a-look-at-what-happens-when-stocks-enter-a-bear-market-2/

Costa, F. J. M. (2017). *Forecasting volatility using GARCH models*. University of Minho.

De Santis, G., & İmrohoroğlu, S. (1997). Stock returns and volatility in emerging financial markets. *Journal of International Money and Finance*, *16*(4), 561–579.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427–431. https://doi.org/10.2307/2286348

Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106. https://doi.org/10.1016/0927-5398(93)90006-D

Duffee, G. R. (1995). Stock returns and volatility a firm-level analysis. *Journal of Financial Economics*, *37*(3), 399–420. https://doi.org/10.1016/0304-405X(94)00801-7

Eloriaga, J. (2018, June). *A Deep Dive on Vector Autoregression in R*. Retrieved on September 10, 2020, from https://towardsdatascience.com/a-deep-dive-on-vector-autoregression-in-r-58767ebb3f06

Enders, W. (2015). Applied Econometric Time Series. (Fourth edition, Vol. 53). John Wiley & Sons, Inc.

Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, *15*(4), 157–168. https://doi.org/10.1257/jep.15.4.157

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, *20*(3), 339–350. https://doi.org/10.1198/073500102288618487

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1007. https://doi.org/10.2307/19127

Engle, R. F., Ito, T., & Lin, W.-L. (1990). Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market. *Econometrica*, *58*(3), 525–542. https://doi.org/10.2307/2938189

Engle, R. F., & Ng, V. K. (1991). Measuring and Testing the Impact of News on Volatility. In *NBER Working Paper Series*. https://doi.org/10.3386/w3681

Engle, R. F., & Sheppard, K. (2001). *THEORETICAL AND EMPIRICAL PROPERTIES OF DYNAMIC CONDITIONAL CORRELATION MULTIVARIATE GARCH*. Retrieved from http://www.nber.org/papers/w8554%0ANATIONAL

Everything you need to know about the 2016 stock market selloff. (2016, January 20). Retrieved on August 31, 2020, from https://www.forbes.com/sites/steveschaefer/2016/01/20/everything-you-need-to-know-about-the-2016-stock-market-selloff/#2d0593083c8e

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, *38*(1), 34–105.

Fisher, L. (1966). Some New Stock-Market Indexes. *The Journal of Business*, *39*(1), 191–225.

Fraser, P. (1996). UK EXCESS SHARE RETURNS: FIRM SIZE AND VOLATILITY. *Scottish Journal of Political Economy*, *43*(1), 71–84. https://doi.org/10.1111/j.1467-9485.1996.tb00839.x

Frazee, G. (2018, December 27). *6 factors that fueled the stock market dive in 2018*. Retrieved on August 30, 2020, from https://www.pbs.org/newshour/economy/making-sense/6-factors-that-fueled-the-stock-market-dive-in-2018

French, K. R., & Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, *17*(1), 5–26. https://doi.org/10.1016/0304-405X(86)90004-8

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, *19*(1), 3–29. https://doi.org/10.1016/0304-405X(87)90026-2

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, *48*(5), 1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(4), 424–438. https://doi.org/10.2307/1912791

Gregoriou, G. N. (2009). Stock Market Volatility. In *Taylor & Francis Group, LLC*. CRC

Press.

Habib, Y., Kiani, Z. I., & Khan, M. A. (2012). Dividend Policy and Share Price Volatility: Evidence from Pakistan. *Global Journal of Management and Business Research*, *12*(5), 79–83.

Harris, R. D. F., & Pisedtasalasai, A. (2006). Return and volatility spillovers between large and small stocks in the UK. *Journal of Business Finance and Accounting*, *33*(9–10), 1556–1571. https://doi.org/10.1111/j.1468-5957.2006.00635.x

Hopkins, H. D. (1988). Firm Size: The Interchangeability of Measures. In *Human Relations* (Vol. 41). Retrieved from http://hjb.sagepub.com.proxy.lib.umich.edu/content/9/2/183.full.pdf+html

Hussainey, K., Mgbame, C. O., & Chijoke-Mgbame, A. M. (2011). Dividend policy and share price volatility: UK evidence. *The Journal of Risk Finance*, *12*(1), 57–68. https://doi.org/10.1108/15265941111100076

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting : Principles and Practice* (2nd Edition).

Isidore, C. (2018, December 31). *2018 was the worst for stocks in 10 years*. Retrieved on August 30, 2020, from https://edition.cnn.com/2018/12/31/investing/dow-stock-market-today/index.html

Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Revie*, *55*(2), 163–172. https://doi.org/10.2307/1403192

Jenkins, J. (2016, January 29). *3 reasons the market has crashed in 2016 – and why the smart money isn't worried*. Retrieved on August 31, 2020, from https://www.fool.com/investing/general/2016/01/29/3-reasons-the-market-has-crashed-in-2016-and-why-t.aspx

Kimberly, J. R. (1976). Organizational Size and the Structuralist Perspective: A Review, Critique, and Proposal. *Administrative Science Quarterly*, *21*(4), 571–597. https://doi.org/10.2307/2391717

Koutmos, G., & Saidi, R. (1995). The Leverage Effect in Individual Stocks and the Debt To Equity Ratio. *Journal of Business Finance & Accounting*, *22*(7), 1063–1075. https://doi.org/10.1111/j.1468-5957.1995.tb00894.x

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, *62*(318), 399–402. https://doi.org/0.2307/2283970

Lin, F. (2019). *VOLATILITY MODELING BASED ON GARCH-SKEWED- T-TYPE MODELS FOR CHINESE STOCK MARKET*. ISCTE Business School.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *75*(2), 297–303.

Long, H. (2016, January 4). Stock market selloff: just how bad is it?. Retrieved on August 31, 2020, from https://money.cnn.com/2016/01/04/investing/stocks-market-dow-selloff/index.html

Maganlal, M. H. (2019). *TAXAS DE CÂMBIO : MODELAÇÃO ARMA-GARCH TAXAS DE CÂMBIO : MODELAÇÃO ARMA-GARCH*. ISCTE-IUL Business School.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, *36*(4), 394–419. https://doi.org/10.1007/978-1-4757-2763-0_14

Mcardle, M. (2020, March 11). *Bear Market Becomes Official with 1,400-Point Drop amid Coronavirus Fears*. Retrieved on August 30, 2020, from https://www.nationalreview.com/news/coronavirus-bear-market-dow-jones-1400-point-drop-amid-coronavirus-fears/

Nazir, M. S., Nawaz, M. M., Anwar, W., & Ahmed, F. (2010). Determinants of stock price volatility in Karachi stock exchange: The mediating role of corporate dividend policy. *International Research Journal of Finance and Economics*, (55), 100–107.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns : A New Approach. *Econometrica*, *59*(2), 347–370. https://doi.org/10.2307/2938260

Nelson, D. B., & Cao, C. Q. (1992). Inequality Constraints in the Univariate GARCH Model. *Journal of Business and Economic Statistics*, *10*(2), 229–235. https://doi.org/10.2307/1391681

Normality Test in R. Retrieved on August 2, 2020, from https://www.datanovia.com/en/lessons/normality-test-in-r/

Officer, R. R. (1973). The Variability of the Market Factor of the New York Stock Exchange. *The Journal of Business*, *46*(3), 434–453.

Orskaug, E. (2009). *Multivariate DCC-GARCH Model-With Various Error Distributions*. Retrieved from https://core.ac.uk/download/pdf/52106361.pdf

Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, *3*(1), 15–102. https://doi.org/10.1016/0927-5398(95)00020-8

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, *55*(2), 277–301. https://doi.org/10.2307/1913237

Phillips, Peter C. B., & Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, *75*(2), 335–346. https://doi.org/10.2307/2336182

Popper, N. & Gough, N. (2015, August 23). *A plunge in China Rattles Markets Across the Globe*. Retrieved on August 29, 2020, from https://www.nytimes.com/2015/08/25/business/dealbook/daily-stock-market-activity.html

Poterba, J. M., & Summers, L. H. (1986). The Persistence of Volatility and Stock Market Fluctuations. *The American Economic Review*, *76*(5), 1142–1151.

Pugh, D. S., Hickson, D. J., Hinings, C. R., & Turner, C. (1969). The Context of Organization Structures. *Administrative Science Quarterly*, *14*(1), 91–114. https://doi.org/10.2307/2391366

Rooney, K. (2018, December 24). *We are now in a bear market- here's what that means*. Retrieved on August 30, 2020, from https://www.cnbc.com/2018/12/24/whats-a-bear-market-and-how-long-do-they-usually-last-.html

Schwert, G. W. (1988). Why Does Stock Market Volatility Change Over Time? In *NBER Working Paper Series* (Vol. 20). Retrieved from http://www.nber.org/papers/w2983

Shalit, S. S., & Sankar, U. (1977). The Measurement of Firm Size. *The Review of Economics and Statistics*, *59*(3), 290–298. https://doi.org/10.2307/1925047

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709

Stekler, H. O. (1964). The Variability of Profitability with Size of Firm, 1947–1958. *Journal of the American Statistical Association*, *59*(308), 1183–1193. https://doi.org/10.2307/2282633

Sweet, K. (2011, August 8). *Dow plunges after S&P downgrade*. Retrieved on August 29, 2020, from https://money.cnn.com/2011/08/08/markets/markets_newyork/index.htm

Taylor, S. (2005). *Asset Price Dynamics, Volatility, and Prediction*. Princeton University Press.

2018: the worst year for stocks since financial crisis. (2019, January 1). Retrieved on August 30, 2020, from https://www.dw.com/en/2018-the-worst-year-for-stocks-since-financial-crisis/a-46915652

Vuolteenaho, T. (2002). What Drives Firm-Level Stock Returns? *The Journal of Finance*, *57*(1), 233–264. https://doi.org/10.1111/1540-6261.00421

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*(5), 931–955. https://doi.org/10.1016/0165-1889(94)90039-6

Extra Bibliography (R documentation)

Trapletti, A., Hornik, K., & LeBaron, B. (2019). *Package 'tseries'*. from https://cran.r-project.org/web/packages/tseries/tseries.pdf

Pfaff, B., Zivot, E., & Stigler, M. (2016). *Package 'urca'*. from https://cran.r-project.org/web/packages/urca/urca.pdf

Qiu, D. (2015). *Package 'aTSA'*. from https://cran.r-project.org/web/packages/aTSA/aTSA.pdf

Hothorn , T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G. & Mitchel, D. (2020). *Package 'lmtest'*. from https://cran.r-project.org/web/packages/lmtest/lmtest.pdf

Pfaff, B., & Stigler, M. (2018). *Package 'vars'*. from https://cran.r-project.org/web/packages/vars/vars.pdf

Ghalanos, A., & Kley, T. (2020). *Package 'rugarch'*. from https://cran.r-project.org/web/packages/rugarch/rugarch.pdf

Ghalanos, A. (2019). *Package 'rmgarch'*. from https://cran.r-project.org/web/packages/rmgarch/rmgarch.pdf

Annexes

Annex A- Augmented Dickey-Fuller Test applied to Russell 2000 Continuously Compounded Returns

Type	Test Statistic	Lags	Significance levels 10pct	5pct	1pct	Value of the test-statistic
None	tau1	6	-1.62	-1.95	-2.58	-18.7324
Drift	tau2	6	-2.57	-2.86	-3.43	-18.773
	phi1		3.78	4.59	6.43	176.2131
Trend	tau3	6	-3.12	-3.41	-3.96	18.7854
	phi2		4.03	4.68	6.09	117.6312
	phi3		5.34	6.25	8.27	176.4463

Annex B- Augmented Dickey-Fuller Test applied to Russell 1000 Continuously Compounded Returns

Type	Test Statistic	Lags	Significance levels 10pct	5pct	1pct	Value of the test-statistic
None	tau1	7	-1.62	-1.95	-2.58	-18.5556
Drift	tau2	7	-2.57	-2.86	-3.43	-18.7106
	phi1		3.78	4.59	6.43	175.0442
Trend	tau3	7	-3.12	-3.41	-3.96	-18.7092
	phi2		4.03	4.68	6.09	116.6802
	phi3		5.34	6.25	8.27	175.0198

Annex C- Phillips-Perron Test applied to Russell 2000 Continuously Compounded Returns

Type	Model	Lags	Significance levels 10pct	5pct	1pct	Value of the test-statistic
Z-tau	Constant	Short	-2.567668	-2.863181	-3.435869	-57.2327
		Long				-57.2098

Annex D- Phillips-Perron Test applied to Russell 1000 Continuously Compounded Returns

Type	Model	Lags	Significance levels			Value of the test-statistic
			10pct	5pct	1pct	
Z-tau	Constant	Short	-2.567668	-2.863181	-3.435869	-58.6214
		Long				-58.7723

Annex E- Kwiatkowski-Phillips-Schmidt-Shin Test applied to Russell 2000 Continuously Compounded Returns

Type	Lags	Significance levels				Value of the test-statistic
		10pct	5pct	2.5pct	1pct	
Tau	Short	0.119	0.146	0.176	0.216	0.0197
	Long					0.02
	Nil					0.0166
Mu	Short	0.347	0.463	0.574	0.739	0.1003
	Long					0.1013
	Nil					0.0848

Annex F- Kwiatkowski-Phillips-Schmidt-Shin Test applied to Russell 1000 Continuously Compounded Returns

Type	Lags	Significance levels				Value of the test-statistic
		10pct	5pct	2.5pct	1pct	
Tau	Short	0.119	0.146	0.176	0.216	0.0197
	Long					0.0208
	Nil					0.0154
Mu	Short	0.347	0.463	0.574	0.739	0.0471
	Long					0.0497
	Nil					0.037

Annex G- Normality tests applied to ARMA model residuals

	Jarque-Bera Test	**Kolmogorov-Smirnov Test**	**Shapiro-Wilk Test**
Russell 2000	< 2.2e-16	< 2.2e-16	< 2.2e-16
Russell 1000	< 2.2e-16	< 2.2e-16	< 2.2e-16

Annex H- Granger causality test

H0: Russell 1000 returns do not Granger-cause the Russell 2000 returns	
F-test	6.0002
df1	3
df2	5056
p-value	0.0004461
H0: Russell 2000 returns do not Granger-cause the Russell 1000 returns	
F-test	4.7206
df1	3
df2	5056
p-value	0.002717

Annex I- Multivariate ARCH test applied to VAR residuals

	ARCH (multivariate)
χ^2	2284.2
df	45
p-value	< 2.2e-16

Annex J- Multivariate Normality tests

	JB Test	**Skewness Test**	**Kurtosis Test**
χ^2	25314	402.57	24911
df	4	2	2
p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16

Annex K- DCC model under normal distribution and student-t distribution (out-of-sample = 150)

		Normal Distribution	Student-t Distribution
GJR-GARCH (1,1)			
Russell 1000	omega	0.000003 (0.000014)	0.000003 (0.352698)
	alpha 1	0.010835 (0.360292)	0.000001 (0.999988)
	beta1	0.815994 (0.000000)	0.815728 (0.000000)
	gamma1	0.234657 (0.000000)	0.290423 (0.000001)
	shape	-	5.321169 (0.000000)
Russell 2000	omega	0.000005 (0.000000)	0.000004 (0.008000)
	alpha 1	0.002004 (0.782550)	0.000004 (0.999649)
	beta1	0.877496 (0.00000)	0.883562 (0.00000)
	gamma1	0.148069 (0.00000)	0.156980 (0.00000)
	shape	-	8.414620 (0.000000)
DCC (1,1)			
	dcca1	0.068155 (0.00000)	0.080511 (0.000000)
	dccb1	0.892191 (0.00000)	0.872552 (0.00000)
	mshape	-	7.620703 (0.000000)
Information Criteria			
AIC		-14.763	-14.827
BIC		-14.722	-14.778

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Annex L- DCC model under normal distribution and student-t distribution (out-of-sample = 250)

		Normal Distribution	Student-t Distribution
GJR-GARCH (1,1)			
Russell 1000	omega	0.000003 (0.000001)	0.000003 (0.480061)
	alpha1	0.007794 (0.493975)	0.000000 (0.999998)
	beta1	0.822221 (0.000000)	0.822113 (0.000000)
	gamma1	0.230804 (0.000000)	0.283224 (0.000137)
	shape	-	5.277542 (0.000000)
Russell 2000	omega	0.000005 (0.000000)	0.000004 (0.000230)
	alpha1	0.003713 (0.631518)	0.000000 (0.999956)
	beta1	0.875509 (0.00000)	0.883143 (0.00000)
	gamma1	0.146926 (0.000000)	0.155332 (0.000000)
	shape	-	8.585897 (0.000000)
DCC (1,1)			
	dcca1	0.065665 (0.00000)	0.075573 (0.00000)
	dccb1	0.898235 (0.00000)	0.882630 (0.00000)
	mshape	-	7.615005 (0.000000)
Information Criteria			
AIC		-14.749	-14.812
BIC		-14.706	-14.762

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Annex M- aDCC model under normal distribution and student-t distribution (out-of-sample = 150)

		Normal Distribution	Student-t Distribution
GJR-GARCH (1,1)			
Russell 1000	omega	0.000003 (0.000014)	0.000003 (0.352626)
	alpha1	0.010835 (0.362775)	0.000001 (0.999988)
	beta1	0.815994 (0.000000)	0.815728 (0.000000)
	gamma1	0.234657 (0.000000)	0.290423 (0.000001)
	shape	-	5.321169 (0.000000)
Russell 2000	omega	0.000005 (0.000000)	0.000004 (0.008292)
	alpha1	0.002004 (0.781082)	0.000004 (0.999648)
	beta1	0.877496 (0.00000)	0.883562 (0.00000)
	gamma1	0.148069 (0.00000)	0.156980 (0.00000)
	shape	-	8.414620 (0.00000)
DCC (1,1)			
	dcca1	0.049014 (0.000023)	0.056756 (0.000001)
	dccb1	0.894912 (0.00000)	0.873501 (0.00000)
	dccg1	0.033099 (0.046089)	0.043580 (0.020483)
	mshape	-	7.942895 (0.000000)
Information Criteria			
	AIC	-14.765	-14.828
	BIC	-14.721	-14.778

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Annex N- aDCC model under normal distribution and student-t distribution (out-of-sample = 250)

		Normal Distribution	Student-t Distribution
GJR-GARCH (1,1)			
Russell 1000	omega	0.000003 (0.000001)	0.000003 (0.480524)
	alpha 1	0.007794 (0.496820)	0.000000 (0.999998)
	beta1	0.822221 (0.000000)	0.822113 (0.000000)
	gamma1	0.230804 (0.00000)	0.283224 (0.000138)
	shape	-	5.277542 (0.000000)
Russell 2000	omega	0.000005 (0.00000)	0.000004 (0.000250)
	alpha 1	0.003713 (0.629226)	0.000000 (0.999956)
	beta1	0.875509 (0.00000)	0.883143 (0.00000)
	gamma1	0.146926 (0.00000)	0.155332 (0.000000)
	shape	-	8.585897 (0.00000)
DCC (1,1)			
	dcca1	0.047239 (0.000025)	0.054071 (0.000002)
	dccb1	0.900539 (0.00000)	0.883788 (0.00000)
	dccg1	0.032144 (0.046012)	0.039184 (0.027452)
	mshape	-	7.937863 (0.000000)
Information Criteria			
AIC		-14.750	-14.814
BIC		-14.705	-14.761

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Annex O- aDCC model under normal distribution and student-t distribution (out-of-sample = 500)

		Normal Distribution	Student-t Distribution
GJR-GARCH (1,1)			
Russell 1000	omega	0.000003 (0.000001)	0.000003 (0.151365)
	alpha 1	0.012058 (0.312877)	0.000002 (0.999949)
	beta1	0.817249 (0.000000)	0.814906 (0.000000)
	gamma1	0.226346 (0.00000)	0.294585 (0.000000)
	shape	-	5.292852 (0.000000)
Russell 2000	omega	0.000005 (0.000000)	0.000004 (0.025704)
	alpha 1	0.005284 (0.511292)	0.000004 (0.999662)
	beta1	0.875671 (0.00000)	0.889367 (0.00000)
	gamma1	0.142289 (0.00001)	0.147700 (0.000001)
	shape	-	8.852236 (0.00000)
DCC (1,1)			
	dcca1	0.048071 (0.000080)	0.054062 (0.000015)
	dccb1	0.898744 (0.00000)	0.884658 (0.00000)
	dccg1	0.029175 (0.083207)	0.037369 (0.046635)
	mshape	-	7.772682 (0.000000)
Information Criteria			
	AIC	-14.787	-14.845
	BIC	-14.738	-14.787

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Annex P- FDCC model under normal distribution and student-t distribution (out-of-sample = 150)

		Normal Distribution
GJR-GARCH (1,1)		
Russell 1000	omega	0.000003 (0.000013)
	alpha 1	0.010835 (0.358400)
	beta1	0.815994 (0.000000)
	gamma1	0.234657 (0.000000)
Russell 2000	omega	0.000005 (0.000000)
	alpha 1	0.002004 (0.781464)
	beta1	0.877496 (0.00000)
	gamma1	0.148069 (0.00000)
DCC (1,1)		
	fdcca1	0.274308 (0.00000)
	fdccb1	0.725692 (0.00000)
Information Criteria		
	AIC	-14.719
	BIC	-14.678

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Annex Q- FDCC model under normal distribution and student-t distribution (out-of-sample = 250)

		Normal Distribution
GJR-GARCH (1,1)		
Russell 1000	omega	0.000003 (0.000001)
	alpha 1	0.007794 (0.492620)
	beta1	0.822221 (0.000000)
	gamma1	0.230804 (0.000000)
Russell 2000	omega	0.000005 (0.000000)
	alpha 1	0.003713 (0.629956)
	beta1	0.875509 (0.00000)
	gamma1	0.146926 (0.00000)
DCC (1,1)		
	fdcca1	0.272844 (0.00000)
	fdccb1	0.727156 (0.00000)
Information Criteria		
AIC		-14.703
BIC		-14.660

Note: the highlighted results correspond to the coefficients that are not statistically significant.

Annex R- FDCC model under normal distribution and student-t distribution (out-of-sample = 500)

		Normal Distribution
GJR-GARCH (1,1)		
Russell 1000	omega	0.000003 (0.000001)
	alpha 1	0.012058 (0.308137)
	beta1	0.817249 (0.000000)
	gamma1	0.226346 (0.000000)
Russell 2000	omega	0.000005 (0.000000)
	alpha 1	0.005284 (0.511473)
	beta1	0.875671 (0.00000)
	gamma1	0.142289 (0.00000)
DCC (1,1)		
	fdcca1	0.264527 (0.00000)
	fdccb1	0.735473 (0.00000)
Information Criteria		
AIC		-14.745
BIC		-14.698

Note: the highlighted results correspond to the coefficients that are not statistically significant.